

Linear-scaling generation of potential energy surfaces using a double incremental expansion

Carolin König^{1, a)} and Ove Christiansen^{1, b)}

Department of Chemistry, Aarhus University, DK-8000 Aarhus C, Denmark.

(Dated: 15 August 2016)

We present a combination of the incremental expansion of potential energy surfaces (PESs), known as n -mode expansion, with the incremental evaluation of the electronic energy in a many-body approach. The application of semi-local coordinates in this context allows the generation of PESs in a very cost-efficient way. For this, we employ the recently introduced flexible adaptation of local coordinates of nuclei (FALCON) coordinates. By introducing an additional transformation step, concerning only a fraction of the vibrational degrees of freedom, we can achieve linear scaling of the *accumulated* cost of the single point calculations required in the PES generation. Numerical examples of these double incremental approaches for oligo-phenyl examples show fast convergence with respect to the maximum number of simultaneously treated fragments and only a modest error introduced by the additional transformation step. The approach, presented here, represents a major step towards the applicability of vibrational wave function methods to sizable, covalently bound systems.

^{a)}Electronic mail: carolink@kth.se; Present address: KTH Royal Institute of Technology, School of Biotechnology, Division of Theoretical Chemistry and Biology, S-106 91 Stockholm, Sweden

^{b)}Electronic mail: ove@chem.au.dk

I. INTRODUCTION

Potential energy surfaces (PESs) are essential ingredients for quantum-dynamical calculations used for the simulation of vibrational spectra and chemical reactions. The focus of the present work is the construction of PESs expanded around a reference point for the use in subsequent anharmonic vibrational wave function calculations. The computational cost of generating full PESs is determined by (i) the number of required single point calculations (SPCs) and (ii) the computational cost per SPC. In a full-dimensional PES generation on a grid, the number of required SPCs increases exponentially with the number of dimensions and the computational cost per SPC scales according to the applied electronic structure method. A full PES generation is, hence, only affordable for molecular systems of a few atoms. A common approximation to overcome the exponential scaling in the number of SPCs is to restrict the direct mode–mode couplings^{1–5} in an incremental-type of expansion. Similarly, the scaling of the electronic structure calculations can be reduced to linear scaling by incremental fragmentation approaches when exploiting the locality of the interaction (see, e.g., Ref. 6). Both of these approaches rely on an incremental expansion and are approximate, but have proven cost effective. In this work, we show, that these two approaches can beneficially be combined. A detailed analysis shows that the computational gains are particularly large when (semi-)local coordinates are employed. By *semi-local* we describe that the majority of the coordinates are strictly local to one or a few fragments. We can even obtain *linear* scaling of the accumulated computational cost of all SPCs required for the PES generation with increasing molecular size. This can be achieved by employing auxiliary vibrational coordinates for certain vibrational degrees of freedom and requires the introduction of an additional transformation step for certain contributions to the PES representation. Notable, this scaling considers the number of SPCs as well as the computational cost of the individual SPCs, and is to our knowledge the first report of such low scaling for anharmonic PES construction considering the full vibrational space.

The above-mentioned restriction of direct mode–mode couplings^{1–5} includes the special case of the pair approximation¹ and is known under different names, such as many-body expansion⁷ n -mode representation³, mode-coupling expansion, cut-HDMR (High Dimensional Model Representation)⁸, cluster expansion⁹, or others¹⁰. We will generally refer to this expansion as n -mode representation or n -mode expansion. This approach reduces the

scaling in required grid points for a PES generation with increasing number of modes from exponential scaling to polynomial scaling with the n th power, where n is the maximum number of simultaneously considered modes. Still, this scaling prevents its application to systems larger than a few tens of vibrational degrees of freedom, especially when combined with decent (and thereby computationally expensive) electronic structure methods. Accordingly, the n -mode representation has been combined with a number of approaches further reducing the number of required SPCs, such as screening schemes^{11–16} and adaptive choice of the grid points^{17–19}. Another possibility to reduce the computational scaling for PES generations is to obtain the expensive higher-order mode couplings in a more approximate manner^{12,20–24}. This can, for example, be achieved by using lower-cost electronic structure methods^{12,20–22}, derivative information²³, or other approximate ways to calculate the individual single point energies²⁴. In special cases, effective linear scaling of the number of required SPCs has been observed with an adaptive grid approach²⁵. Both screening as well as adapted choice of required grid points are expected to be particularly beneficial, when combined with adapted coordinates with some kind of spatial locality^{26–29}. Still, even if linear scaling of the number of grid points can be achieved (which is likely to be system and coordinate dependent for the described approaches), the computational cost of every grid point still scales according to the chosen electronic structure method with system size. This typically leads to overall super-linear scaling with system size.

Incremental and fragmentation ideas for electronic calculations have been found in the literature for several decades now and have raised considerable attention within the last decades (see the historical review in Ref. 30). This is reflected in a large number of different incremental fragmentation approaches^{6,31–35}. Examples are the fragment molecular orbital (FMO)^{31,32,36} model, the molecular fractioning with conjugated caps (MFCC) scheme³³, the systematic molecular fragmentation (SMF) method⁶, and the generalized energy-based fragmentation (GEFB) method³⁵, just to name a few. The fragmentation idea has also been combined with multi-layer methods.^{37–43} A full account of all fragmentation methods is far beyond the scope of the present introduction. We refer to a recent review focusing on FMO methods⁴⁴ as well as comprehensive reviews on energy-based fragmentation methods^{30,45} and other recent attempts to classify the different methods⁴² and/or generalize the underlying expressions and ideas to a unified framework^{46–48}.

Our double incremental approach concerns a combination of the incremental expansion of

the PES in terms of vibrational degrees of freedom (known as n -mode representation) with that of the electronic energy in terms of fragments in a flexible manner. We are not aware of any previous work achieving such a flexible setup. This makes it fundamentally different from the *many-body expansion* approach for PES generation of Varandas and Murrell⁴⁹ and its so-called *double many-body expansion*⁵⁰ extension, despite the similar names. In the present work, particular gains are obtained by employing fairly local vibrational coordinates. The attractiveness of employing "local" modes in combination with the n -mode representation is also exploited in the local monomer model by Bowman and co-workers^{51–53}. Indeed, one of the main characteristics of the local monomer model is that all coordinates are fully localized to a particular fragment and as such it has been very successfully applied in the context of PES generation for molecular clusters. Our setup is not restricted to local vibrational coordinates only. The presented double incremental scheme can thereby be applied to all types of vibrational coordinates and particular savings are obtained for all, though significant savings are most easily achievable for coordinates that are strictly local to part of the system. Our methods do, however, not require that all coordinates are local. Thereby our approach is applicable to covalently bound systems. Still, to achieve a significant reduction of the computational cost of the presented double incremental approach compared to conventional approaches, semi-local coordinates should be applied. In the present work, we make use of the recently introduced FALCON (Flexible Adaptation of Local COordinates of Nuclei) scheme⁵⁴, which aims at constructing exactly such coordinate sets. The resulting rectilinear FALCON coordinates are constructed in a way that a large fraction of the vibrational coordinates are local to certain groups of atoms (i.e., local to fragments) and therefore enable a spatially motivated incremental expansion of the PES. Using these coordinates in our double incremental expansion will be denoted *double incremental expansion in FALCON coordinates* (DIF). In this model, the part of the FALCON coordinates that are not fully local can lead to a super-linear scaling of the number of required SPCs. This is why, we have devised an additional protocol, which generates local auxiliary coordinates covering the same vibrational space as the common FALCON coordinates for a local combination of fragments. These auxiliary coordinates are generally different for different fragments and combinations of fragments. Performing the SPCs for the PES construction along these auxiliary coordinates allows linear scaling in the accumulated cost of SPCs. It, however, requires the introduction of an additional transformation

step for the local PES representation to be used in an overall PES expansion. We term this approach *double incremental expansion in FALCON coordinates with auxiliary coordinate transformation* (DIFACT). This transformation of parts of the PES is, however, only exact in the limit of a complete PES and therefore introduces an additional potential source of error in practical calculations. This lifts the exact limits of the incremental expansion in case of DIFACT and must be numerically validated.

In this article, we derive and motivate the DIF as well as the DIFACT model starting from general incremental expansion including scaling considerations (Section II). We furthermore present a pilot implementation (Section III), computational setup (Section IV) as well as first numerical results for tetra- and hexa-phenyl (Section V). After concluding from our results, we give a perspective on future extensions and applications of the presented method (Section VI).

II. THEORY

A. General incremental expansion

In this section, we will set up a common nomenclature for incremental expansions, which is capable of describing the n -mode expansion of PESs and the incremental expansion of the energy (often denoted by many-body expansion⁵⁵) at a given point using the same principles and nomenclature. As outlined in the Introduction both expansions are well-known in literature and the purpose of the present section is to reformulate the basic ideas in a general and convenient form. Thus, deviations from standard formulations found in literature are not fundamental, but necessary to formulate the double incremental expansion in a compact and general manner in Sections II B – II G.

The main idea of incremental expansions is that a multi-dimensional function $F(x_1, x_2, \dots, x_n)$ can be expanded around a reference point $F^0 = F(0, 0, \dots, 0)$ with in-

dices running from 1 to n as

$$\begin{aligned}
F(x_1, x_2, \dots, x_n) = & F^0 + \sum_{i=1}^n [F_i^{1D}(x_i) - F^0] \\
& + \sum_{j>i=1}^n \{ [F_{ij}^{2D}(x_i, x_j) - F^0] - [F_i^{1D}(x_i) - F^0] - [F_j^{1D}(x_j) - F^0] \} \\
& + \sum_{k>j>i=1}^n \{ [F_{ijk}^{3D}(x_i, x_j, x_k) - F^0] \\
& \quad - [F_{ij}^{2D}(x_i, x_j) - F^0] - [F_i^{1D}(x_i) - F^0] - [F_j^{1D}(x_j) - F^0] \\
& \quad - [F_{ik}^{2D}(x_i, x_k) - F^0] - [F_i^{1D}(x_i) - F^0] - [F_k^{1D}(x_k) - F^0] \\
& \quad - [F_{jk}^{2D}(x_j, x_k) - F^0] - [F_j^{1D}(x_j) - F^0] - [F_k^{1D}(x_k) - F^0] \\
& \quad - [F_i^{1D}(x_i) - F^0] - [F_j^{1D}(x_j) - F^0] - [F_k^{1D}(x_k) - F^0] \} \\
& + \dots,
\end{aligned} \tag{1}$$

where the following lower-order *cut functions* are used,

$$F^0 \equiv F(0, 0, \dots, 0, 0, 0, \dots, 0, 0, 0, \dots, 0, 0, 0, \dots, 0) \tag{2}$$

$$F_i^{1D}(x_i) \equiv F(0, 0, \dots, 0, x_i, 0, \dots, 0, 0, 0, \dots, 0, 0, 0, \dots, 0) \tag{3}$$

$$F_{ij}^{2D}(x_i, x_j) \equiv F(0, 0, \dots, 0, x_i, 0, \dots, 0, x_j, 0, \dots, 0, 0, 0, \dots, 0) \tag{4}$$

$$F_{ijk}^{3D}(x_i, x_j, x_k) \equiv F(0, 0, \dots, 0, x_i, 0, \dots, 0, x_j, 0, \dots, 0, x_k, 0, \dots, 0). \tag{5}$$

This expansion is exact in the limit where the order of the expansion is equal to the number of variables.

Introducing the bar functions,

$$\bar{F}^{0D} \equiv F^0 \tag{6}$$

$$\bar{F}_i^{1D}(x_i) \equiv F_i^{1D}(x_i) - \bar{F}^{0D} \tag{7}$$

$$\bar{F}_{ij}^{2D}(x_i, x_j) \equiv F_{ij}^{2D}(x_i, x_j) - \bar{F}_i^{1D}(x_i) - \bar{F}_j^{1D}(x_j) - \bar{F}^{0D} \tag{8}$$

$$\begin{aligned}
\bar{F}_{ijk}^{3D}(x_i, x_j, x_k) \equiv & F_{ijk}^{3D}(x_i, x_j, x_k) - \bar{F}_{ij}^{2D}(x_i, x_j) - \bar{F}_{ik}^{2D}(x_i, x_k) - \bar{F}_{jk}^{2D}(x_j, x_k) \\
& - \bar{F}_i^{1D}(x_i) - \bar{F}_j^{1D}(x_j) - \bar{F}_k^{1D}(x_k) - \bar{F}^{0D}
\end{aligned} \tag{9}$$

\vdots

we can rewrite Eq. (1) as

$$F(x_1, x_2, \dots, x_n) = \bar{F}^{0D} + \sum_{i=1}^n \bar{F}_i^{1D}(x_i) + \sum_{j>i=1}^n \bar{F}_{ij}^{2D}(x_i, x_j) + \sum_{k>j>i=1}^n \bar{F}_{ijk}^{3D}(x_i, x_j, x_k) + \dots \quad (10)$$

In analogy to the set logic applied in Ref. 56, we define a variable combination range (VCR) containing all variable combinations that are explicitly parameterized in the expansion. Variable combinations are simply sets of variables and denoted in bold, such as \mathbf{c}_l . It is often convenient to include l in the notation, where l is the number of variables in this set. The VCR is required to be *closed on forming subsets*. A VCR is *closed on forming subsets*, if for each variable combination \mathbf{c}_l in the VCR, also all subsets of \mathbf{c}_l are considered. This definition includes the empty set in all cases. For instance, for a three-variable function $F(i, j, k)$ that is treated to second order in the incremental expansion, the variable combination range is $\{\{\}, \{i\}, \{j\}, \{k\}, \{i, j\}, \{i, k\}, \{j, k\}\}$.

We can now rewrite Eq. (10) in a compact and flexible form as

$$F(x_1, x_2, \dots, x_n) = \sum_{\mathbf{c}_l \in \{\text{VCR}\}} \bar{F}^{\mathbf{c}_l}(\{x\}^{\mathbf{c}_l}), \quad (11)$$

where the equal sign only holds for a complete VCR. $\bar{F}^{\mathbf{c}_l}(\{x\}^{\mathbf{c}_l})$ is the bar function for the variable combination \mathbf{c}_l , i.e., a function of l variables. The particular set of variables given by \mathbf{c}_l is extracted from the full set, as indicated by the notation $\{x\}^{\mathbf{c}_l}$. For instance for $\mathbf{c}_l = \mathbf{c}_3 = \{i, j, k\}$, we have $\bar{F}^{\{i,j,k\}}(\{x\}^{\{i,j,k\}}) = \bar{F}_{ijk}^{3D}(x_i, x_j, x_k)$.

Using the introduced notation we express the bar potentials generally as

$$\bar{F}^{\mathbf{c}_l}(\{x\}^{\mathbf{c}_l}) = F^{\mathbf{c}_l}(\{x\}^{\mathbf{c}_l}) - \sum_{\substack{\mathbf{c}_s \subset \mathbf{c}_l \\ \mathbf{c}_s \in \{\text{VCR}\}}} \bar{F}^{\mathbf{c}_s}(\{x\}^{\mathbf{c}_s}). \quad (12)$$

This formulation not only allows for more compact expressions, but also enables a more flexible parameterization of the expansion, such as opening for different treatment of variables with different degree of correlation. We accordingly restrict the sum in Eq. (12) to those lower-order fragment combinations present in the VCR. This restriction is not needed whenever the VCR is closed on forming subsets (as assumed by default). We note, that with this restriction, Eq. (12) also holds in case of effective VCRs, which are introduced in Section II B.

In the following, we need to express the bar functions for a variable combination in terms of the original, uncorrected contributions, i.e., cut functions. Most generally, we can write

$$\bar{F}^{\mathbf{c}_l}(\{x\}^{\mathbf{c}_l}) = \sum_{\substack{\mathbf{c}_{l'} \subseteq \mathbf{c}_l \\ \mathbf{c}_{l'} \in \{\text{VCR}\}}} k^{\mathbf{c}_{l'}, \mathbf{c}_l} F^{\mathbf{c}_{l'}}(\{x\}^{\mathbf{c}_{l'}}), \quad (13)$$

where $k^{\mathbf{c}_{l'}, \mathbf{c}_l}$ is the coefficient of $F^{\mathbf{c}_{l'}}(\{x\}^{\mathbf{c}_{l'}})$ in $\bar{F}^{\mathbf{c}_l}(\{x\}^{\mathbf{c}_l})$. As a consequence of the closed under forming subset condition of the VCR, all lower-order variable combinations are contained in the VCR. In this case, the coefficient amounts to $k^{\mathbf{c}_{l'}, \mathbf{c}_l} = (-1)^{l-l'}$ (see also Ref. 56).

B. Effective variable combination ranges

As stated before, we generally require the VCR to be closed on forming subsets. Depending on the chosen VCR, it can, however, happen that the coefficients of the cut functions for some variable combinations are zero in the overall expansion. Accordingly, these variable combinations are omitted in the corresponding *effective* VCR. These cases occur in VCRs in which not all variable couplings are included up to the same coupling level. To identify these terms, we first write the full function as a weighted sum of its cut functions, i.e., we combine Eqs. (11) and (13) to

$$F(\{x\}) = \sum_{\mathbf{c}_l \in \{\text{VCR}\}} \sum_{\substack{\mathbf{c}_{l'} \subseteq \mathbf{c}_l \\ \mathbf{c}_{l'} \in \{\text{VCR}\}}} (-1)^{l-l'} F^{\mathbf{c}_{l'}}(\{x\}^{\mathbf{c}_{l'}}) = \sum_{\mathbf{c}_{l'} \in \{\text{VCR}\}} p_{\mathbf{c}_{l'}}^{\text{VCR}} F^{\mathbf{c}_{l'}}(\{x\}^{\mathbf{c}_{l'}}) \quad (14)$$

with

$$p_{\mathbf{c}_{l'}}^{\text{VCR}} = \sum_{\mathbf{c}_l \in \text{VCR}; \mathbf{c}_l \supseteq \mathbf{c}_{l'}} (-1)^{l-l'} = \sum_{l=l'}^{l_{\max}} N_{\mathbf{c}_{l'}, l}^{\text{VCR}} (-1)^{l-l'}, \quad (15)$$

where l_{\max} is the largest variable-combination order in the VCR and $N_{\mathbf{c}_{l'}, l}^{\text{VCR}}$ is the number of variable combinations of the order l that are supersets of or equal to $\mathbf{c}_{l'}$ and are contained in the VCR, which is closed on forming subsets. For a more rigorous derivation of Eq. (15) we refer to Appendix A. We can, hence, use Eq. (15) to analyze the incremental expansion for a given VCR and set up an *effective* VCR, in which all variable combinations with zero contributions are omitted. Effective VCRs are not generally closed on forming subsets. An incremental expansion with such an effective VCR, will, however, give the same results as the analogous expansion in the corresponding VCR that contains all subsets of variable

combinations. When setting up an effective VCR, special care should be taken, keeping in mind that the extension of the VCR may lead to other previously omitted terms to have a non-zero contribution.

Assume, we expand the function $F(x_A, x_B, x_C, x_D)$ in a variable combination range $\text{VCR}_1 = \{\{\}, \{A\}, \{B\}, \{C\}, \{D\}, \{A, B\}, \{C, D\}\}$ which is closed on forming subsets. We can calculate the coefficient for the variable combination $\{A\}$ in an incremental expansion for VCR_1 by

$$p_{\{A\}}^{\text{VCR}_1} = N_{\{A\},1}^{\text{VCR}_1}(-1)^{1-\dim(\{A\})} + N_{\{A\},2}^{\text{VCR}_1}(-1)^{2-\dim(\{A\})}, \quad (16)$$

where $N_{\{A\},1}^{\text{VCR}_1} = 1$ is the number of variable combinations in VCR_1 of order 1 that are supersets of or equal to $\{A\}$. This only holds for the variable combination $\{A\}$ itself. Similarly, for $N_{\{A\},2}^{\text{VCR}_1}$, we count the number of variable combinations in VCR_1 of order 2 that are supersets of or equal to $\{A\}$. This only holds true for $\{A, B\}$, so that $N_{\{A\},2}^{\text{VCR}_1} = 1$. Inserting $N_{\{A\},1}^{\text{VCR}_1} = N_{\{A\},2}^{\text{VCR}_1} = 1$ into Eq. (16), we obtain $p_{\{A\}}^{\text{VCR}_1} = 0$. The corresponding coefficients for $\{B\}$, $\{C\}$, and $\{D\}$ are obtained in complete analogy to that for $\{A\}$ for the given VCR_1 . This means that the variable combinations $\{A\}$, $\{B\}$, $\{C\}$, and $\{D\}$ have zero contribution in the incremental expansion for VCR_1 . We can, hence, construct an effective VCR for VCR_1 that contains only variable combinations with non-zero coefficients $p_{\mathbf{c}_i'}^{\text{VCR}_1}$. It reads $\text{VCR}_1^{\text{eff}} = \{\{\}, \{A, B\}, \{C, D\}\}$.

Let us now consider the very similar $\text{VCR}_2 = \{\{\}, \{A\}, \{B\}, \{C\}, \{D\}, \{A, B\}, \{A, C\}, \{C, D\}\}$. It differs from VCR_1 only by the addition the variable combination $\{A, C\}$. The individual coefficients can be calculated with an equation that is similar to Eq. (16), but replaces VCR_1 by VCR_2 . Counting the variable combinations in VCR_2 that are supersets of or equal to $\{A\}$ in the variable combination orders 1 and 2, separately, we obtain $N_{\{A\},1}^{\text{VCR}_2} = 1$ and $N_{\{A\},2}^{\text{VCR}_2} = 2$. The two variable combinations of order 2 in VCR_2 that contain $\{A\}$ are $\{A, B\}$ and $\{A, C\}$. Employing VCR_2 , the variable combination $\{A\}$ has, hence, the coefficient $p_{\{A\}}^{\text{VCR}_2} = -1$. The same coefficient is obtained for $\{C\}$. The number of variable combinations of order 2 in VCR_2 that are supersets of $\{B\}$ is $N_{\{B\},2}^{\text{VCR}_2} = 1$. Together with $N_{\{B\},1}^{\text{VCR}_2} = 1$, we obtain $p_{\{B\}}^{\text{VCR}_2} = 0$. $p_{\{D\}}^{\text{VCR}_2} = 0$ is obtained completely analogously. This means, that for VCR_2 , only $\{B\}$ and $\{D\}$ and also the empty set $\{\}$ (not shown) have zero contribution. With these coefficients we can now assemble an effective VCR for VCR_2 with only non-zero contributions. It reads $\text{VCR}_2^{\text{eff}} = \{\{A\}, \{C\}, \{A, B\}, \{A, C\}, \{C, D\}\}$.

Obviously, the variable combinations $\{A\}$ and $\{C\}$, which can be omitted in $\text{VCR}_1^{\text{eff}}$, cannot be omitted in $\text{VCR}_2^{\text{eff}}$, even though the only difference between VCR_1 and VCR_2 is the additional variable combination $\{A, C\}$. We will come back to effective VCRs, when analyzing the incremental energy expressions for calculating total energies of a molecular system from fragments and fragment combinations (see Section II D 2).

C. Uncoupled variables

In the derivation of the double incremental scheme with semi-local coordinates, we will meet cases, where variables of a function are uncoupled. In the following, we will give an induction argument, why the bar functions, containing uncoupled variables are zero. We denote the uncoupled variable y and let l be the number of other variables from which y is uncoupled. Throughout this section, we assume the VCR to be closed on forming subsets.

Considering first the $l = 1$ case, we denote the two variables by x and y and call them uncoupled in F if

$$F(x, y) = F(x, 0) + F(0, y) - F(0, 0). \quad (17)$$

The corresponding bar functions can be expanded according to Eqs. (12) and (13). We obtain $\bar{F}_1(x) = F(x, 0) - F(0, 0)$ and $\bar{F}_2(y) = F(0, y) - F(0, 0)$ for the 1D functions. For the 2D function we find that $\bar{F}(x, y) = F(x, y) - \bar{F}_1(x) - \bar{F}_2(y) - F^0 = F(x, 0) + F(0, y) - F(0, 0) - F(x, 0) + F(0, 0) - F(0, y) + F(0, 0) - F(0, 0) = 0$. Thus, the two-dimensional bar function with two uncoupled variables is exactly zero.

We assume the induction hypothesis $\bar{F}(\{x\}^{\mathbf{c}_s}, y) = 0$ for $s = l - 1$ with $s, l \in \mathbb{N}$. We then set out to show that this also implies that $\bar{F}(\{x\}^{\mathbf{c}_l}, y) = 0$. Again, y is uncoupled from a set $\{x\}^{\mathbf{c}_l}$ of other coordinates, and we can write, similar to Eq. (17),

$$F(\{x\}^{\mathbf{c}_l}, y) = F(\{x\}^{\mathbf{c}_l}, 0) + F(\{0\}^{\mathbf{c}_l}, y) - F(\{0\}^{\mathbf{c}_l}, 0). \quad (18)$$

Introducing Eq. (18) into Eq. (12) and separating the VCR summation in three terms, we obtain

$$\begin{aligned} \bar{F}^{\mathbf{c}_l, y}(\{x\}^{\mathbf{c}_l}, y) = & F(\{x\}^{\mathbf{c}_l}, 0) + F(\{0\}^{\mathbf{c}_l}, y) - F(\{0\}^{\mathbf{c}_l}, 0) \\ & - \sum_{\mathbf{c}_s \subseteq \mathbf{c}_l} \bar{F}^{\mathbf{c}_s}(\{x\}^{\mathbf{c}_s}) - \sum_{\mathbf{c}_s \subset \mathbf{c}_l, \mathbf{c}_s \neq \emptyset} \bar{F}^{\mathbf{c}_s, y}(\{x\}^{\mathbf{c}_s}, y) - \bar{F}(y). \end{aligned} \quad (19)$$

Exploiting that the value of $F(\{x\}^{\mathbf{c}_l}, 0)$ is equal to that of $F(\{x\}^{\mathbf{c}_l}) = \sum_{\mathbf{c}_s \subseteq \mathbf{c}_l} \bar{F}^{\mathbf{c}_s}(\{x\}^{\mathbf{c}_s})$ for a given $\{x\}^{\mathbf{c}_l}$ and $\bar{F}(y) = F(\{0\}^{\mathbf{c}_l}, y) - F(\{0\}^{\mathbf{c}_l}, 0)$, one sees that the only contributions to the bar function of a variable combination with an uncoupled variable arise from the lower-level variable combinations with one uncoupled variable,

$$\bar{F}^{\mathbf{c}_l, y}(\{x\}^{\mathbf{c}_l}, y) = - \sum_{\mathbf{c}_s \subseteq \mathbf{c}_l, \mathbf{c}_s \neq \emptyset} \bar{F}^{\mathbf{c}_s, y}(\{x\}^{\mathbf{c}_s}, y), \quad (20)$$

which is zero by the induction assumption. One can, hence, conclude by induction that this also holds for any variable combination that contains a variable that is uncoupled to the others.

D. Application of the incremental expansion to potential energy surfaces and evaluation of electronic energies

1. Incremental expansion of the potential energy surface

The application of this framework to multi-dimensional PESs is equivalent to the well-known n -mode expansions of the PESs outlined in the Introduction. The PES is expressed in our notation (similar to that in Ref. 56) as,

$$V(\{q\}) \approx \sum_{\mathbf{m}_n \in \{\text{MCR}\}} \bar{V}^{\mathbf{m}_n}(\{q\}^{\mathbf{m}_n}) = \sum_{\mathbf{m}_n \in \{\text{MCR}\}} \sum_{\substack{\mathbf{m}_{n'} \subseteq \mathbf{m}_n \\ \mathbf{m}_{n'} \in \{\text{MCR}\}}} k^{\mathbf{m}_{n'}, \mathbf{m}_n} V^{\mathbf{m}_{n'}}(\{q\}^{\mathbf{m}_{n'}}). \quad (21)$$

The coordinates $\{q\}$ can describe all possible arrangements of the nuclei. They may be in principle any set of such non-redundant coordinates. \mathbf{m}_n is a composite index labeling a mode combination, which is a set of n modes. All mode combinations that are explicitly parameterized in a given expansion of $V(\{q\})$ are then collected in the so-called mode combination range (MCR), which is accordingly the set of mode combinations. $\{q\}^{\mathbf{m}_n}$ is the subset of all n vibrational coordinates in $\{q\}$ that are contained in the mode combination \mathbf{m}_n . The energy origin can be chosen freely, we assume $V^0 = 0$ at the expansion point. Accordingly, the applied nomenclature is in complete analogy to the more general one introduced above.

2. Incremental expansion of the electronic energy in fragmented approaches

The electronic energy of a system consisting of fragments can be expressed in an incremental expansion in terms of its fragments. This has often been denoted by many-body expansion. We can formulate this expansion of the electronic energy similar to the one derived for PESs in Section II D 1: Assuming a system built up from N fragments, we define $E = E(z_1, z_2, \dots, z_N)$ as the energy of the total system. z_1 , z_2 , and z_N are composite indices representing particular conformations of the fragments 1, 2, and N , respectively. The zero values of these composite indices $z_i = 0$ are defined such that in this case no atoms of fragment i are present in the respective conformation. We further define the energy of a fragment combination in the conformation $\{z\}_{\mathbf{f}_i}$ as $E_{\mathbf{f}_i}(\{z\}_{\mathbf{f}_i})$ and set the expansion point to $E_0 = 0$. Again, we employ the above-introduced nomenclature, but now the variables describe the conformation of the individual fragments. This means we expand the total energy in the framework described above by setting up a *fragment combination range* (FCR) containing all considered fragment combinations \mathbf{f}_i . It reads [analogously to Eq. (11)]

$$E(z_1, z_2, \dots, z_N) \approx \sum_{\mathbf{f}_i \in \{\text{FCR}\}} E_{\mathbf{f}_i}(\{z\}_{\mathbf{f}_i}), \quad (22)$$

where

$$E_{\mathbf{f}_i}(\{z\}_{\mathbf{f}_i}) = E_{\mathbf{f}_i}(\{z\}_{\mathbf{f}_i}) - \sum_{\substack{\mathbf{f}_{i'} \subset \mathbf{f}_i \\ \mathbf{f}_{i'} \in \{\text{FCR}\}}} \underline{E}_{\mathbf{f}_{i'}}(\{z\}_{\mathbf{f}_{i'}}) \quad (23)$$

is the contribution of the fragment combination \mathbf{f}_i to the overall energy and is evaluated analogously to Eq. (12). This expression is equivalent to the incremental formulations being very widespread in quantum chemistry, as outlined in the Introduction. For the fragment combinations we use subscripts instead of superscripts for consistency to Sections II E – II G. Note, that we also use underbars instead of bars to indicate corrections with respect to lower-order fragment combinations.

Eqs. (22) and (23) represent a very general energy expression for incremental molecular fragmentation approaches: The energy expressions for different standard approaches are obtained for different choices of FCR. This formulation is thereby related to other attempts to generalize the many-body expansion such as the generalized many-body (GMB) expansion

by Richard and Herbert⁴⁶ or the many-overlapping-body (MOB) expansion by Mayhall and Raghavachari⁴⁷. The starting points of the different approaches are rather different in the sense that both the GMB and MOB expansions consider overlapping fragments, whereas we always have disjoint fragments as smallest units. These are combined to larger, overlapping fragment combinations in the incremental expansion. A detailed comparison of these formulations and full account of the different fragmentation schemes covered clearly goes beyond the scope of the present article. Still, we would like to reason that our scheme, which starts from disjoint fragments, can describe standard fragmentation methods with overlapping fragments as well. Here, we use similar arguments as discussed in Ref. 30. The energy expression considering overlapping fragments often contains contributions for overlaps of fragments. Assume we have a system ABC, to which we assign overlapping fragments AB and BC. A typical approach in fragmentation schemes with overlapping fragments would then be to add the energies of AB and BC and subtract the energy of the overlap, i.e., that of B. The energy for this example is, hence, calculated as $E \approx E_{AB} + E_{BC} - E_{AB \cap BC} = E_{AB} + E_{BC} - E_B$. In our framework, we start out from non-overlapping fragments, i.e., A, B, and C and then set up a FCR, which is closed on forming subsets. Such an FCR could, for example read $\text{FCR} = \{\{\}, \{A\}, \{B\}, \{C\}, \{A, B\}, \{B, C\}\}$. Employing Eq. (15), we can determine the coefficients of all contributions of the individual terms, i.e., of the cut functions. This reveals, that the only non-zero contributions in this case are $p_{\{A, B\}}^{\text{FCR}} = p_{\{B, C\}}^{\text{FCR}} = 1$ and $p_{\{B\}}^{\text{FCR}} = -1$, so that the overall energy expression is $E \approx E_{AB} + E_{BC} - E_B$ and thereby equivalent to that obtained with overlapping fragments. We anticipate, that the translation of the fragmentation approaches with overlapping fragments to that presented here can generally be obtained by (i) defining disjoint fragments from the overlaps of the fragments, (ii) identifying the largest fragment combinations to be considered in the expansion, (iii) setting up a FCR that is closed on forming subsets, and (iv) calculating the coefficients for each fragment combination according to Eq. (15) and thereby reduce the FCR, that is closed on forming subsets, to an effective FCR.

In this sense, incremental fragmentation approaches using overlapping fragments correspond to the incremental energy expansion used here with spatial constraints to the considered coupling terms. In our terminology, this corresponds to a truncation of the FCR with spatial considerations. If we, for example, consider a chain like system ABCD, the FCR up

to three-fragment interactions reads

$$\begin{aligned} \text{FCR}^{3\text{F}} = \{ \{\}, \{A\}, \{B\}, \{C\}, \{D\}, \{A, B\}, \{A, C\}, \{A, D\}, \{B, C\}, \{B, D\}, \{C, D\}, \\ \{A, B, C\}, \{A, B, D\}, \{A, C, D\}, \{B, C, D\} \}. \end{aligned} \quad (24)$$

Restricting this FCR to direct neighbor coupling, we obtain

$$\begin{aligned} \text{FCR}^{3\text{F}, \text{NB}} = \{ \{\}, \{A\}, \{B\}, \{C\}, \{D\}, \{A, B\}, \{A, C\}, \{B, C\}, \{B, D\}, \{C, D\}, \\ \{A, B, C\}, \{B, C, D\} \}, \end{aligned} \quad (25)$$

where $\{A, C\}$ and $\{B, D\}$ are not direct neighbors, but have to be included to ensure that the FCR is closed on forming subsets. Such spatial restrictions of the FCR will be essential to obtain linear scaling approaches, since each fragment has a limited number of neighbors. In many cases of these spatially restricted FCRs, it will be possible to omit certain lower-order fragment combinations in the expansion based on Eq. (15), see also Appendix B. Computational savings due to the exploitation of such effective FCRs, will be comparably small, since such considerations only allow to neglect lower-order fragment combinations and, thus, do not affect the leading terms in the computational cost. The latter arise from the highest-order fragment combinations in the FCR. See also the more detailed discussion on the scaling behavior of the different double incremental schemes for generating PESs in Section II H.

3. Translation of a point on the potential energy surface to the energy of a given conformation

So far, we have expressed the PES as well as the individual energy points in incremental expansions. Note, that we have employed two different sets of coordinates. These are $\{z\}$, defining the actual conformation, and $\{q\}$, a set of generalized internal vibrational coordinates. The latter specifies the displacement from a reference structure. Since we can express both by the other, i.e., write $\mathbf{z} = \mathbf{z}(\mathbf{q})$ and $\mathbf{q} = \mathbf{q}(\mathbf{z})$, these are interchangeable representations. In case of rectilinear coordinates, we can express the positions (\mathbf{z}_i) of all n nuclei in terms of nuclear positions for a reference configuration (\mathbf{r}_0) and rectilinear displacements (\mathbf{d}_i) as $\mathbf{z}_i = \mathbf{r}_0 + \mathbf{d}_i$. These displacements can then be collected in a displacement vector $\mathbf{d}^T = (d_{1x}, d_{1y}, d_{1z}, d_{2x}, d_{2y}, d_{2z}, \dots, d_{nx}, d_{ny}, d_{nz})$. Rectilinear vibrational coordinates can

then be obtained by orthogonal transformation of the mass-scaled Cartesian displacement vectors ($\mathbf{d}_m = \mathbf{M}^{\frac{1}{2}}\mathbf{d}$) by $\mathbf{q} = \mathbf{L}^T\mathbf{d}_m$. \mathbf{M} is a diagonal matrix of the size $3n \times 3n$ set up from n diagonal 3×3 matrices of the type $m_i\mathbf{1}$, where m_i is the mass of the respective nucleus. With different orthogonal transformation matrices \mathbf{L} , where $\mathbf{L}^T\mathbf{L} = \mathbf{1}$, different sets of coordinates can be constructed. The column vectors of \mathbf{L} , denoted by \mathbf{l}_k , for $k = 1, \dots, 3n$, define the corresponding coordinates as

$$q_k = \mathbf{l}_k^T \mathbf{d}_m. \quad (26)$$

Accordingly, a displacement along vibrational coordinates \mathbf{q} is directly related to a certain configuration of the system by

$$\mathbf{z} = \mathbf{r}_0 + \mathbf{M}^{-\frac{1}{2}}\mathbf{L}\mathbf{q}, \quad (27)$$

defining a clear transformation between these variables. In the discussion of the FALCON coordinates and PES transformations below, we will use the \mathbf{L} matrices and their column vectors \mathbf{l}_k to describe coordinates, keeping in mind that Eq. (26) provides the rigorous connection.

E. Double incremental expansion of the potential energy surface

In the following, the above incremental expansions of the PES and the energy will be combined in a general manner, i.e., for any MCR and FCR satisfying the condition to be closed on forming subsets. To establish a connection between the PES points and the energy evaluations, we first define the value of a sub PES for the mode combination \mathbf{m}_n at a grid point $\{q\}^{\mathbf{m}_n}$ by the difference of its electronic energy to the energy of the reference structure $E(\mathbf{r}_0)$,

$$V^{\mathbf{m}_n}(\{q\}^{\mathbf{m}_n}) = E(\{z([q]^{\mathbf{m}_n})\}) - E(\mathbf{r}_0) = \Delta E(\{z([q]^{\mathbf{m}_n})\}), \quad (28)$$

where $[q]^{\mathbf{m}_n} = \{\{q\}^{\mathbf{m}_n}, \{0\}^{m \notin \mathbf{m}_n}\}$ is a shorthand notation for a full set of modes where only those in \mathbf{m}_n can be different from zero.

We then insert an incremental expansion for the energy of the displaced as well as reference structure, as given in Eq. (22), employing the same FCR for both incremental energy evaluations,

$$V^{\mathbf{m}_n}(\{q\}^{\mathbf{m}_n}) \approx \sum_{\mathbf{f}_l \in \{\text{FCR}\}} [E_{\mathbf{f}_l}(\{z([q]^{\mathbf{m}_n})\}_{\mathbf{f}_l}) - E_{\mathbf{f}_l}(\mathbf{r}_{0,\mathbf{f}_l})] = \sum_{\mathbf{f}_l \in \{\text{FCR}\}} \Delta E_{\mathbf{f}_l}(\{z([q]^{\mathbf{m}_n})\}_{\mathbf{f}_l}) \quad (29)$$

where \mathbf{f}_l is a fragment combination of size l in the FCR and the sum runs over all fragment combinations in the FCR. $\mathbf{r}_{0,\mathbf{f}_l}$ denotes the reference conformation for the fragment combination \mathbf{f}_l . The corrected energy for a given fragment combination can then be expressed by using Eq. (13) as

$$\Delta E_{\mathbf{f}_l}(\{z([q]^{\mathbf{m}_n})\}_{\mathbf{f}_l}) = \sum_{\substack{\mathbf{f}_{l'} \subseteq \mathbf{f}_l \\ \mathbf{f}_{l'} \in \{\text{FCR}\}}} k_{\mathbf{f}_{l'}, \mathbf{f}_l} \Delta E_{\mathbf{f}_{l'}}(\{z([q]^{\mathbf{m}_n})\}_{\mathbf{f}_{l'}}). \quad (30)$$

We can then combine the above-obtained expressions Eq. (30), Eq. (29), and Eq. (21) to the double incremental expression for the PES,

$$V(\{q\}) \approx \sum_{\mathbf{m}_n \in \{\text{MCR}\}} \sum_{\substack{\mathbf{m}_{n'} \subseteq \mathbf{m}_n \\ \mathbf{m}_{n'} \in \{\text{MCR}\}}} k^{\mathbf{m}_{n'}, \mathbf{m}_n} \sum_{\mathbf{f}_l \in \{\text{FCR}\}} \sum_{\substack{\mathbf{f}_{l'} \subseteq \mathbf{f}_l \\ \mathbf{f}_{l'} \in \{\text{FCR}\}}} k_{\mathbf{f}_{l'}, \mathbf{f}_l} \Delta E_{\mathbf{f}_{l'}}(\{z([q]^{\mathbf{m}_{n'}})\}_{\mathbf{f}_{l'}}). \quad (31)$$

For later usage, we introduce the shorthand notation

$$\Delta E_{\mathbf{f}_{l'}}^{\mathbf{m}_{n'}} \equiv \Delta E_{\mathbf{f}_{l'}}(\{z([q]^{\mathbf{m}_{n'}})\}_{\mathbf{f}_{l'}}). \quad (32)$$

With this, we write the general double incremental expansion of the PES as,

$$V(\{q\}) \approx \sum_{\mathbf{m}_n \in \{\text{MCR}\}} \sum_{\substack{\mathbf{m}_{n'} \subseteq \mathbf{m}_n \\ \mathbf{m}_{n'} \in \{\text{MCR}\}}} k^{\mathbf{m}_{n'}, \mathbf{m}_n} \sum_{\mathbf{f}_l \in \{\text{FCR}\}} \sum_{\substack{\mathbf{f}_{l'} \subseteq \mathbf{f}_l \\ \mathbf{f}_{l'} \in \{\text{FCR}\}}} k_{\mathbf{f}_{l'}, \mathbf{f}_l} \Delta E_{\mathbf{f}_{l'}}^{\mathbf{m}_{n'}}. \quad (33)$$

The individual contributions to this expansion are, hence, energy differences for a given mode combination ($\mathbf{m}_{n'}$) in a given fragment combination ($\mathbf{f}_{l'}$). So far, we have interpreted the double incremental expansion as an n -mode expansion with incremental energy evaluations. We will see in the following, that breaking the PES contributions down to the basic terms of $\Delta E_{\mathbf{f}_{l'}}^{\mathbf{m}_{n'}}$ will give us the flexibility to reformulate the double incremental expansion as an incremental expansion of the PES built up from PESs of contributing fragment combinations. In the following, it will be shown, that this interpretation in combination with semi-local coordinates will enable the generation of full PESs with beneficial scaling.

F. Semi-local coordinates

To achieve beneficial computational scaling with the double incremental scheme, we need to use semi-local coordinates as outlined below. The defining characteristic of such coordinate sets is that some of the coordinates are constrained to internal motions in a limited set

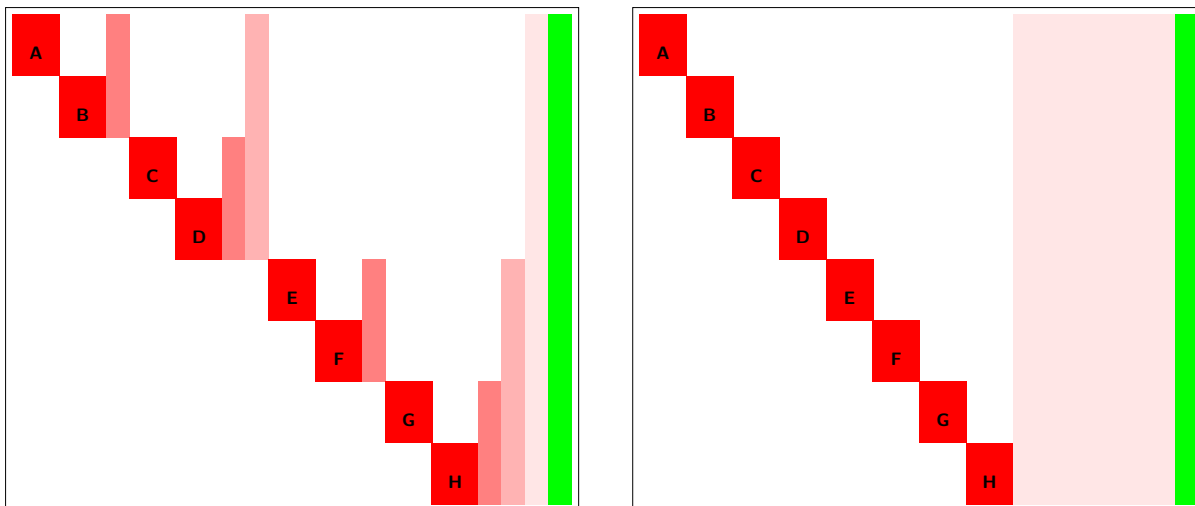


FIG. 1. Spatial character of FALCON coordinates for two examples with eight fragments (A, B, C, D, E, F, G, and H) with semi-local (left) and delocalized (right) inter-connecting modes. The squares symbolize the \mathbf{L} matrices for these FALCON coordinates. These matrices only have non-zero entries on the colored areas. The red blocks symbolize semi-local vibrational coordinates, where the deepest red indicates the largest degree of localization and the green areas correspond to the overall translational and rotational coordinates.

of atoms and are thereby only *active* for a limited number of fragments and *rigid* in others. The term rigid is used to underline that the *internal* structure within these fragments stays unchanged by the displacement, but the *relative orientation* to other fragments may be varied. The discussion below on the usage of semi-local coordinates in the double incremental expansion of PESs holds for all types of coordinates fulfilling these criteria, regardless of the exact details on how these coordinates are set up.

The set of semi-local coordinates used in this study are the recently introduced FALCON coordinates⁵⁴. The FALCON algorithm enables the generation of a full, rectilinear set of purely vibrational coordinates with well-defined spatial character. Possible structures of the \mathbf{L} matrix for such FALCON coordinates are sketched in Figure 1. Some of the FALCON coordinates only have contributions from a well-defined group of atoms, i.e., they are localized to a certain fragment (indicated by the deepest red in Figure 1). All other, so-called *inter-connecting* coordinates span more than one fragment and are depicted in lighter red in Figure 1. Depending on the settings of the FALCON scheme, the inter-connecting

coordinates might all span the entire molecule or only a few fragments. The total number of these inter-connecting coordinates is, however, in all cases smaller than or equal to $(F - 1)6$, where F is the total number of fragments. Their number is, thus, significantly smaller than the total number of modes.

The inter-connecting coordinates are special kinds of coordinates in the sense that the fragments or groups of fragments move relative to each other as *semi-rigid* groups. We choose the term semi-rigid, since the inter-connecting coordinates arise from the local translational and rotational degrees of freedom of the groups, and are hence rigid in the infinitesimal description. In a rectilinear space, as applied in the FALCON scheme, however, the relative rotation of different groups cannot be described as rotations of internally fully rigid groups for larger displacements. This is why these types of coordinates are not strictly rigid.

We denote the inter-connecting coordinates that move the semi-rigid entities A and B relative to each other by $(A) \leftrightarrow (B)$ and those that move the fragment pairs (AB) and (CD) as semi-rigid entities by $(AB) \leftrightarrow (CD)$. Occasionally, we also need to account for the truly rigid fragments in the formulation. We mark these fragments and fragment combinations by square brackets. Thereby we can differentiate between the inter-connecting modes for the left and right examples, respectively, in Figure 1. In the left case, the inter-connecting modes, i.e., $(A) \leftrightarrow (B) - [CDEFGH]$, $[AB] - (C) \leftrightarrow (D) - [EFGH]$, $[ABCD] - (E) \leftrightarrow (F) - [GH]$, $[ABCDEF] - (G) \leftrightarrow (H)$, $(AB) \leftrightarrow (CD) - [EFGH]$, $[ABCD] - (EF) \leftrightarrow (GH)$, and $(ABCD) \leftrightarrow (EFGH)$, have distinct and different spatial structures. All inter-connecting modes in the right case are of type $(A) \leftrightarrow (B) \leftrightarrow (C) \leftrightarrow (D) \leftrightarrow (E) \leftrightarrow (F) \leftrightarrow (G) \leftrightarrow (H)$ and thereby span the entire system.

All considered FALCON coordinates diagonalize a reduced mass-weighted Hessian matrix for a subspace of the vibrational space of the entire system. By this procedure, we can assign *quasi-harmonic* frequencies to these FALCON coordinates calculated from the diagonal elements of the reduced problem in the usual manner. For a more detailed description of the character of FALCON coordinates, we refer to the section II B in Ref. 54.

G. Double incremental expansion in semi-local coordinates

The double incremental PES expansion, Eq. (33), has been derived and can be interpreted as an n -mode expansion with incremental energy evaluation. Turning the order of summation

in the double incremental expansion around,

$$V(\{q\}) \approx \sum_{\mathbf{f}_l \in \{\text{FCR}\}} \sum_{\substack{\mathbf{f}_{l'} \subseteq \mathbf{f}_l \\ \mathbf{f}_{l'} \in \{\text{FCR}\}}} k_{\mathbf{f}_{l'}, \mathbf{f}_l} \sum_{\mathbf{m}_n \in \{\text{MCR}\}} \sum_{\substack{\mathbf{m}_{n'} \subseteq \mathbf{m}_n \\ \mathbf{m}_{n'} \in \{\text{MCR}\}}} k^{\mathbf{m}_{n'}, \mathbf{m}_n} \Delta E_{\mathbf{f}_{l'}}^{\mathbf{m}_{n'}}, \quad (34)$$

we see that we can equally well interpret this expansion as an incremental expansion of PESs of molecular fragments. Therefore, we introduce the PES for the fragment combination \mathbf{f}_l , which can be expanded in an n -mode expansion as

$$V_{\mathbf{f}_l}(\{q\}) \approx \sum_{\mathbf{m}_n \in \{\text{MCR}\}} \sum_{\substack{\mathbf{m}_{n'} \subseteq \mathbf{m}_n \\ \mathbf{m}_{n'} \in \{\text{MCR}\}}} k^{\mathbf{m}_{n'}, \mathbf{m}_n} \Delta E_{\mathbf{f}_l}^{\mathbf{m}_{n'}} = \sum_{\mathbf{m}_n \in \{\text{MCR}\}} \bar{V}_{\mathbf{f}_l}^{\mathbf{m}_n}(\{q\}^{\mathbf{m}_n}), \quad (35)$$

where $\bar{V}_{\mathbf{f}_l}^{\mathbf{m}_n}(\{q\}^{\mathbf{m}_n})$ is the corresponding bar potential for the mode combination \mathbf{m}_n . In this interpretation, it is striking that the PES of every fragment combination is spanned by the same number of modes as the total system. This number is significantly higher than the number of degrees of freedom in the respective fragment combination.

Applying semi-local coordinates, however, all coordinates that are rigid within the respective fragment combination are uncoupled from the other modes. These have, hence, zero contribution to the respective bar potentials $\bar{V}_{\mathbf{f}_l}^{\mathbf{m}_n}(\{q\}^{\mathbf{m}_n})$, see also Section II C. We define $\mathbf{m}_{r, \mathbf{f}_l}$ as a set of modes containing all r modes that are rigid in the fragment combination \mathbf{f}_l , i.e., do not change the internal structure of \mathbf{f}_l . With this, we can restrict the MCR for the PES of the fragment combination \mathbf{f}_l ,

$$V_{\mathbf{f}_l}(\{q\}) = \tilde{V}_{\mathbf{f}_l}(\{q\} \setminus \{q\}^{\mathbf{m}_{r, \mathbf{f}_l}}) \approx \sum_{\substack{\mathbf{m}_n \in \{\text{MCR}\} \\ \mathbf{m}_n \cap \mathbf{m}_{r, \mathbf{f}_l} = \emptyset}} \bar{V}_{\mathbf{f}_l}^{\mathbf{m}_n}(\{q\}^{\mathbf{m}_n}), \quad (36)$$

where the MCR has to be closed on forming subsets. Here, the tilde on the V is simply to mark that this function is formally different from the standard V functions since it is a function of fewer coordinates, but with the same values. Notice that in Eq. (36) only modes contribute that are non-rigid in the respective fragment combination. We can now rewrite the full PES, Eq. (34), as

$$V(\{q\}) \approx \sum_{\mathbf{f}_l \in \{\text{FCR}\}} \sum_{\substack{\mathbf{f}_{l'} \subseteq \mathbf{f}_l \\ \mathbf{f}_{l'} \in \{\text{FCR}\}}} k_{\mathbf{f}_{l'}, \mathbf{f}_l} \tilde{V}_{\mathbf{f}_{l'}}(\{q\} \setminus \{q\}^{\mathbf{m}_{r, \mathbf{f}_{l'}}}) = \sum_{\mathbf{f}_l \in \{\text{FCR}\}} \tilde{V}_{\mathbf{f}_l}(\{q\} \setminus \{q\}^{\mathbf{m}_{r, \mathbf{f}_l}}), \quad (37)$$

This expansion of the PES is particularly appealing since it significantly reduces the number of modes to be considered in the PESs for the individual fragment combinations, when dealing with semi-local modes. Thereby the computational cost of generating the PESs can be significantly reduced as outlined in detail in Section II H.

H. Concrete computational methods and computational scaling

The computational cost of a PES generation is governed by the accumulated computational costs of the required SPCs, which is determined by the number of individual SPCs needed and the cost per SPC. Unfortunately, as the system size increases, both of these factors increase drastically. We will see that the double incremental expansion proposed in this work can reduce the computational cost of PES generations and the associated scaling with increasing system size. For a static grid, we can directly calculate the number of required SPCs for the different approaches. We assume that all of the fragments are non-linear and therefore have six nuclear degrees of freedom from translational and rotational motions. We further assume throughout this section a chain-like system and account only for direct neighbor couplings. The assumed FALCON coordinates are of the same type as those shown in the right part of Figure 1. In this way, we obtain $(F - 1)6$ completely delocalized interconnecting coordinates, where F is the total number of fragments. We furthermore exploit the reduction of required terms by forming an effective FCR for these kinds of systems (see Appendix B). In the present case, only the largest fragment combinations contribute as well as the second largest ones that do not contain an outer-most fragment of the chain-like system.

In this section, we will develop the number of required SPCs and formal computational scaling for the different expansions of the PES assuming the particular setup of a chain-like system described above. This is a significant simplification and many real systems will be far from this simple. However, this simplification is convenient for the clarity of the argument. It is expected that even if the simplest nearest neighbor description is not fully adequate for all systems, most systems will be such that a given fragment has only significant interaction with a limited set of other fragments. As long as this limited set of fragments does not increase with the size of the system the analysis is still relevant for a rough scaling estimate, though not intended to be a precise prediction of the actual computational cost. The derived benefits for the double incremental schemes compared to the conventional models are, hence, expected to be valid for a wide range of systems.

In the following equations, N_{pfr} stands for the number of atoms per fragment, which is assumed to be the same for every fragment, n is the maximal number of modes per mode combination, and f is the maximal number of fragments per fragment combination. We

use g for the number of grid points per mode and s as scaling exponent for the underlying electronic structure method. This means, the cost of the electronic structure calculations is assumed to be proportional to $(N_{\text{at}})^s$ for a system with N_{at} atoms. For a given fragment combination of l fragments, the computational cost will thus be proportional to $(l \cdot N_{\text{pfr}})^s$. Assuming N_{pfr} is constant, the interesting aspect is that the computational cost for the most demanding SPC in a double incremental PES generation increases as f^s with increasing maximal fragment combination level f .

1. *Conventional n -mode expansion*

For the conventional n -mode expansion the number of SPCs is given by

$$N_{\text{SPC}}^{n\text{-mode}} = \sum_{m=0}^n \binom{F(3N_{\text{pfr}} - 6) + (F - 1)6}{m} g^m, \quad (38)$$

where the total number of vibrational modes ($M_{\text{tot}} = 3FN_{\text{pfr}} - 6$) is calculated as F times the number of modes per fragment ($3N_{\text{pfr}} - 6$) plus the remaining part of $(F - 1)6$ modes. The number of mode combinations of mode-combination level m is then obtained as the binomial coefficient of $\binom{M_{\text{tot}}}{m}$. Each of these mode combinations requires g^m SPCs. In the limit of a large number of fragments, the number of SPCs scales with F^n . And every individual SPC exhibits an F^s scaling. The combined scaling is given as F^{n+s} .

2. *Double incremental expansion in normal modes (DIN)*

Using the same n -mode expansion as in Section IIH1 but with incremental energy calculations, the evaluation of the energy difference for a particular conformation requires one SPC for each fragment combination that has a non-zero contribution to the overall energy. This is the case for all fragment combinations in normal modes, since a normal mode will generally alter the internal structure of every fragment in the system. For the given setup of a chain-like system with neighbor coupling and a maximal fragment combination level of f , we need to consider $(F - f + 1)$ fragment combinations of size f as well as $(F - f)$ fragment combinations of size $(f - 1)$. The contributions of all lower-order fragment combinations are exactly zero in this setup as outlined in Appendix B. The number $(F - f)$ is obtained from the total number of $(F - (f - 1) + 1)$ connected fragment combinations of size $(f - 1)$ minus

2. By this, we account for the fact, that the two outer-most fragment combinations have zero-contribution (see Appendix B). We can then calculate the overall number of required SPCs as

$$N_{\text{SPC}}^{\text{DIN}} = \left[\sum_{m=0}^n \binom{F(3N_{\text{pfr}} - 6) + (F - 1)6}{m} g^m \right] \cdot \left[\sum_{l=f-1}^f (F - l + (-1)^{f-l}) \right]. \quad (39)$$

We use the squared brackets to highlight that the number of grid points in the overall PES (left bracket) is independent of the number of fragment combinations in the FCR (right bracket). The latter number can be interpreted as the number of SPCs per grid point of the full system. An alternative viewpoint on these two terms in brackets would be that the left bracket contains the number of grid points or SPCs required to generate the PES of one fragment combination. This number is then multiplied with the number of fragment combinations. In any case, the overall scaling in the number of SPCs for large F is F^{n+1} . The gain here lies in the fact that the cost per SPC is largely reduced and is independent of F . Accordingly, the scaling of the accumulated computational cost of all SPCs is likewise F^{n+1} . This scaling behavior is not limited to the double incremental scheme in normal modes. In fact, it holds for all types of coordinates that are not strictly rigid in the sense that all modes can alter the internal structure of all fragments and fragment combinations.

3. *Double incremental expansion in FALCON coordinates (DIF)*

When using FALCON coordinates, we can restrict the number of modes in the PESs for the individual fragment combinations to those that are non-rigid in the respective fragment combination, as shown in Section II G. Thereby we can reduce the number of modes to be considered in a PES construction for a fragment combination of size l to $l(3N_{\text{pfr}} - 6) + (F - 1)6$, where we assume that all $(F - 1)6$ inter-connecting modes contribute. The overall number of required SPCs is then reduced to

$$N_{\text{SPC}}^{\text{DIF}} = \sum_{l=f-1}^f (F - l + (-1)^{f-l}) \sum_{m=0}^n \left[\binom{l(3N_{\text{pfr}} - 6) + (F - 1)6}{m} g^m \right]. \quad (40)$$

Note, that we cannot turn these sums freely around since the number of SPCs required in the PES generation of the respective fragment combination depends on the fragment combination. Despite the drastic reduction of the number of required SPCs, we obtain the

same overall formal scaling of F^{n+1} as in Section II H 2. This is due to the fact that the maximal number of modes in a fragment combination still generally scales linear with F . We shall see shortly that DIF should be expected to provide a much-reduced computational effort compared to a standard approach. Nevertheless, we proceed to consider reducing the formal computational scaling as well.

4. *Double incremental expansion in FALCON coordinates with auxiliary coordinate transformation (DIFACT)*

In Section II H 2, the PES of a fragment combination is calculated in the same number of modes as the complete system. The number of relevant modes for a fragment combination has then been drastically reduced by introducing semi-local FALCON coordinate in Section II H 3. We now go one step further and introduce auxiliary inter-connecting modes, spanning the vibrational space of the inter-connecting modes within each fragment combination in a non-redundant manner. For a fragment combination consisting of l fragments, we have, next to the $l(3N_{\text{pfr}} - 6)$ intra-fragment modes, $(l - 1)6$ vibrational degrees of freedom, which we express as inter-connecting modes. In rectilinear coordinates, we have to include additional three coordinates describing the overall infinitesimal rotation of the fragment combination. This is due to the limitations of rectilinear vibrational coordinates with respect to the description of “real” rotations for non-infinitesimal displacements. The overall number of required SPCs is then obtained as

$$N_{\text{SPC}}^{\text{DIFACT}} = \sum_{l=f-1}^f (F - l + (-1)^{f-l}) \sum_{m=0}^n \left[\binom{l(3N_{\text{pfr}} - 6) + (l - 1)6 + 3}{m} g^m \right]. \quad (41)$$

In this case, the number of considered modes for a certain fragment combination is determined by the size of this fragment combination and independent of F , and so is the number of SPCs required to generate the PES of the respective fragment combination. The number of fragment combinations to be considered, however, still scales linear with F , and the total number of required SPCs scales likewise. Since also here the cost per individual SPC is independent of F , the linear scaling is also obtained when considering the accumulated cost for all SPCs that are required when constructing the PES in a the double incremental manner. The introduction of specific auxiliary coordinates for the generation of the PESs for individual fragment combinations, has, however, the disadvantage, that it requires subsequent

transformation of the PESs to the common, overall coordinates. Such a transformation is only exact in case of a full PES representation. Since this is typically not feasible, we will generally introduce a transformation error when going from DIF to DIFACT. The implemented transformation step exhibits polynomial scaling with the number of modes (see Section II J for further details). Its overall cost is, however, typically orders of magnitudes lower than the explicit calculation of the corresponding SPCs, and the computational cost of the transformation has therefore so far not been a bottleneck.

5. *Example scaling behavior*

The number of SPCs for the different schemes with increasing number of fragments is shown in Figure 2 for a chain-like system assuming fragments of ten atoms (24 modes per fragment) and eight grid points per mode. In these examples, we assume the exact same type of system as in the scaling considerations above. It is seen that the variation in the actual number of SPCs matches well the asymptotic predictions of $\log(\#SP) \approx \log(F)$ and $\log(\#SP) \approx (n + 1) \log(F)$ for DIFACT and DIN or DIF, respectively, where $n = 2$ in the present example. The number of required SPCs in the DIN and DIF schemes lies above the number of SPCs required for the conventional treatment for all system sizes. For the DIFACT scheme, this only holds for a small number of fragments. In this scheme, the number of required SPCs increases less drastically with the number of fragments than for the conventional n -mode treatment. The crossover points between the number of SPCs in DIFACT with that for the conventional n -mode expansion are at system sizes around four fragments for $f = 2$ and eleven fragments for $f = 3$.

The analysis so far is limited by the fact that the number of SPCs is not representative for the total computational cost, since the individual SPCs can have quite different cost and they may refer to SPCs with different number of atoms. To estimate the computational cost for a PES generation, we have weighted each SPC by the factor $A \cdot (N_{\text{at,FC}})^s$ mimicking the effect of the different computational cost. $N_{\text{at,FC}}$ is the number of atoms in the relevant fragment combination. The resulting estimates for $s = 1, 3, 5, 7$ with $n = 2$ and $f = 2$ are plotted in the four lower graphs in Figure 2. These exponents have been chosen to represent the computational scaling of electronic structure methods in standard implementations, i.e., third order for density functional theory (DFT), fifth order for second-

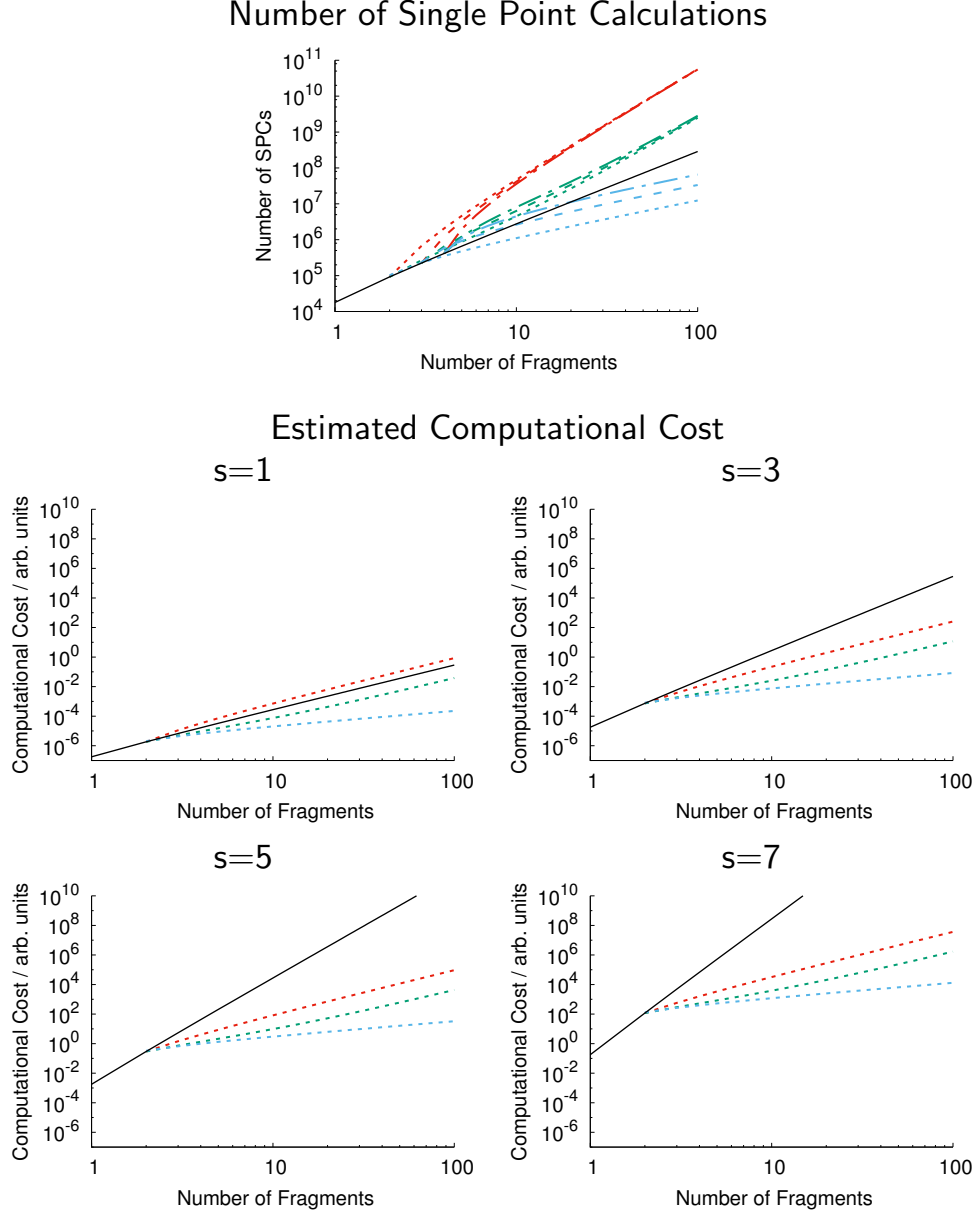


FIG. 2. Upper part: Number of single point calculations (SPCs) for PES generations with the different schemes, i.e., conventional n -mode representation (black), DIN (red), DIF (green), and DIFACT (blue) according to Eqs. (38), (39), (40), and (41), respectively, for different number of fragments ($F = 1 - 100$) and different fragment-combination level, i.e., $f = 2$ (- - -), $f = 3$ (---), and $f = 4$ (- · - ·) in the double incremental schemes distinguished by the line type. The default values are $N_{\text{pfr}} = 10$, $n = 2$, $g = 8$. Bottom part: Estimated computational cost obtained for $f = 2$ by multiplying the number of SPCs for a certain system size of $l \cdot N_{\text{pfr}}$ atoms by $A \cdot (l \cdot N_{\text{pfr}})^s$, where $s = 1, 3, 5, 7$ and $A = 10^{-12}$. It employs the same color code for the different methods as the upper part.

order Møller–Plesset (MP2) calculations and seventh order for coupled cluster with doubles and perturbative triples [CCSD(T)]. Additionally, linear scaling approaches are considered. Linear scaling methods have been a major focus of electronic structure development work in the last decades, and we have here included it to show clearly that even in this case the expected computational cost for the double incremental approach in FALCON coordinates is lower than for the conventional n -mode expansion combined with a linear scaling algorithm for the SPCs. This also holds for the DIF scheme, which requires more SPCs than the conventional n -mode expansion. The observation that the overall cost increases, when combining the DIN scheme with linear scaling approaches for the SPCs, reflects that the application of the fragmentation approach to already linear scaling electronic structure methods cannot reduce the scaling further, but introduces extra costs. This combination is not a typical application at any means.

Besides a general increase in computational cost with increasing s , we also see that the double incremental schemes get more and more beneficial for larger values of s . For systems with more than a few fragments we find orders of magnitude efficiency gains in DIN, DIF, and DIFACT methods compared to the conventional n -mode expansion. Since the individual SPCs have similar cost in DIN, DIF, and DIFACT, the additional order of magnitude gains from DIN to DIF and DIFACT derives from the reduction in the number of SPCs seen in the upper part of the figure. Not surprisingly, the additional computational cost, when increasing the maximal fragment combination level, is larger the higher the scaling for the applied electronic structure method is, i.e., the larger s (see Figure S-1 in the supplementary material⁵⁷).

I. Auxiliary coordinates

As seen in Section II H 4, we can obtain linear scaling of the accumulated cost of the required SPCs in a PES generation, as long as the number of modes to be accounted for in the generations of the PESs for the individual fragment combinations does not scale with the overall number of fragments. Such a situation might be achievable in internal, curvilinear modes. In rectilinear coordinates, which allow easy evaluation of the kinetic energy in vibrational structure calculations, we are not aware of local purely vibrational coordinates that fulfill this requirement generally. We propose the usage of alternative auxiliary coordi-

nates for the generation of the PESs of the individual fragment combinations. These PESs in auxiliary coordinates need to be transformed to the respective PES in the common coordinates before the linear combination of the PES for the individual fragment combinations is possible. This transformation is in practice approximate (see Sections II J and III C). The resulting transformation error can be expected to be smaller the more the final coordinates are resembled by the auxiliary coordinates. This means that it is desirable to obtain the auxiliary coordinates in a similar manner as the common inter-connecting coordinates. In the present case, we use FALCON-type coordinates for the common coordinates, which suggests the use of a similar scheme for the generation of the auxiliary coordinates. We, construct the auxiliary coordinates as FALCON-type inter-connecting coordinates that are local to the respective fragment combination. For a more detailed description of the applied protocol, we refer to Section III A.

J. Transformation of potential energy surfaces of fragment combinations from auxiliary to common FALCON coordinates

With the PES transformation, we aim at a PES representation for a fragment combination (FC) in the local redundant coordinates that corresponds to the common FALCON coordinates ($\mathbf{L}_{\text{FC}}^{\text{F}}$). The starting point of the transformation is a PES representation, in which some of the vibrational degrees of freedom are expressed in auxiliary coordinates $\mathbf{L}_{\text{FC}}^{\text{aux}}$. Let us first assume, our set of auxiliary coordinates includes vibrational inter-connecting (I), as well as rotational (R) and translational (T) coordinates of the fragment combination,

$$\mathbf{L}_{\text{FC}}^{\text{aux}} = (\mathbf{L}_{\text{FC}}^{\text{aux}}|^{\text{I}} \mathbf{L}_{\text{FC}}^{\text{aux}}|^{\text{R}} \mathbf{L}_{\text{FC}}^{\text{aux}}|^{\text{T}}). \quad (42)$$

We refer to Section III A for a more detailed discussion on how these coordinates are obtained. For now, we only need to know that the column vectors of $\mathbf{L}_{\text{FC}}^{\text{aux}}$ are orthonormal. In this case, the complete space spanned by $\mathbf{L}_{\text{FC}}^{\text{F}}$ is also contained in $\mathbf{L}_{\text{FC}}^{\text{aux}}$ and we can write

$$\mathbf{L}_{\text{FC}}^{\text{F}} = \mathbf{L}_{\text{FC}}^{\text{aux}} \mathbf{A}_{\text{FC}}, \quad (43)$$

where the rectangular transformation matrix is set up as

$$\mathbf{A}_{\text{FC}} = (\mathbf{L}_{\text{FC}}^{\text{aux}})^T \mathbf{L}_{\text{FC}}^{\text{F}} = \begin{pmatrix} \mathbf{L}_{\text{FC}}^{\text{aux}}|^{\text{I}} \mathbf{L}_{\text{FC}}^{\text{F}} \\ \mathbf{L}_{\text{FC}}^{\text{aux}}|^{\text{R}} \mathbf{L}_{\text{FC}}^{\text{F}} \\ \mathbf{L}_{\text{FC}}^{\text{aux}}|^{\text{T}} \mathbf{L}_{\text{FC}}^{\text{F}} \end{pmatrix}, \quad (44)$$

so that

$$\mathbf{L}_{\text{FC}}^{\text{F}} = {}^{\text{I}}\mathbf{L}_{\text{FC}}^{\text{aux}} {}^{\text{I}}\mathbf{A}_{\text{FC}} + {}^{\text{R}}\mathbf{L}_{\text{FC}}^{\text{aux}} {}^{\text{R}}\mathbf{A}_{\text{FC}} + {}^{\text{T}}\mathbf{L}_{\text{FC}}^{\text{aux}} {}^{\text{T}}\mathbf{A}_{\text{FC}}. \quad (45)$$

The molecular structure of a fragment combination for a displacement along these common coordinates around the reference point $(\mathbf{r}_0)_{\text{FC}}$ can be obtained using Eq. (27) as

$$\begin{aligned} \mathbf{z}_{\text{FC}}^{\text{F}} &= (\mathbf{r}_0)_{\text{FC}} + \mathbf{M}_{\text{FC}}^{-\frac{1}{2}} \mathbf{L}_{\text{FC}}^{\text{F}} \mathbf{q}_{\text{FC}}^{\text{F}} \\ &= (\mathbf{r}_0)_{\text{FC}} + \mathbf{M}_{\text{FC}}^{-\frac{1}{2}} {}^{\text{I}}\mathbf{L}_{\text{FC}}^{\text{aux}} {}^{\text{I}}\mathbf{A}_{\text{FC}} \mathbf{q}_{\text{FC}}^{\text{F}} + \mathbf{M}_{\text{FC}}^{-\frac{1}{2}} {}^{\text{R}}\mathbf{L}_{\text{FC}}^{\text{aux}} {}^{\text{R}}\mathbf{A}_{\text{FC}} \mathbf{q}_{\text{FC}}^{\text{F}} + \mathbf{M}_{\text{FC}}^{-\frac{1}{2}} {}^{\text{T}}\mathbf{L}_{\text{FC}}^{\text{aux}} {}^{\text{T}}\mathbf{A}_{\text{FC}} \mathbf{q}_{\text{FC}}^{\text{F}} \\ &= (\mathbf{r}_0)_{\text{FC}} + \Delta^{\text{I}}\mathbf{r}_{\text{FC}}^{\text{F}} + \Delta^{\text{R}}\mathbf{r}_{\text{FC}}^{\text{F}} + \Delta^{\text{T}}\mathbf{r}_{\text{FC}}^{\text{F}}. \end{aligned} \quad (46)$$

The subscript FC indicates that only the contributions of the atoms in a particular fragment combination are considered, e.g., $\mathbf{q}_{\text{FC}}^{\text{F}}$ denotes the part of the common FALCON coordinates that effect the fragment combination FC. $\Delta^{\text{I}}\mathbf{r}_{\text{FC}}^{\text{F}}$, $\Delta^{\text{R}}\mathbf{r}_{\text{FC}}^{\text{F}}$, and $\Delta^{\text{T}}\mathbf{r}_{\text{FC}}^{\text{F}}$ are the displacements within this fragment combination along the common FALCON coordinates due to vibrational inter-connecting, rotational, and translational motions, respectively. The energy of a fragment combination FC for a particular displacement along $\mathbf{L}_{\text{FC}}^{\text{F}}$, can, hence, be expressed as

$$\begin{aligned} E_{\text{FC}}(\{z(\{q[\mathbf{L}_{\text{FC}}^{\text{F}}]\})\}_{\text{FC}}) &= E_{\text{FC}}((\mathbf{r}_0)_{\text{FC}} + \Delta^{\text{I}}\mathbf{r}_{\text{FC}}^{\text{F}} + \Delta^{\text{R}}\mathbf{r}_{\text{FC}}^{\text{F}} + \Delta^{\text{T}}\mathbf{r}_{\text{FC}}^{\text{F}}) \\ &= E_{\text{FC}}((\mathbf{r}_0)_{\text{FC}} + \Delta^{\text{I}}\mathbf{r}_{\text{FC}}^{\text{F}} + \Delta^{\text{R}}\mathbf{r}_{\text{FC}}^{\text{F}}) \\ &= E_{\text{FC}}(\{z(\{q[{}^{\text{I}}\mathbf{L}_{\text{FC}}^{\text{aux}} {}^{\text{I}}\mathbf{A}_{\text{FC}} + {}^{\text{R}}\mathbf{L}_{\text{FC}}^{\text{aux}} {}^{\text{R}}\mathbf{A}_{\text{FC}}]\})\}_{\text{FC}}), \end{aligned} \quad (47)$$

where $\{q[\mathbf{L}]\}$ expresses that the set of coordinates $\{q\}$ is defined through the \mathbf{L} matrix. The second equality in Eq. (47) holds, due to the translational invariance of the energy. As previously noted, the rotational coordinates will not generally be purely rotational, and we cannot do a similar trick for these as long as we are using rectilinear coordinates. Since all points on the PES are calculated from the electronic energy points, similar arguments can be used to express

$$V_{\text{FC}}(\{q[\mathbf{L}_{\text{FC}}^{\text{F}}]\}) = V_{\text{FC}}(\{q[{}^{\text{I}}\mathbf{L}_{\text{FC}}^{\text{aux}} {}^{\text{I}}\mathbf{A}_{\text{FC}} + {}^{\text{R}}\mathbf{L}_{\text{FC}}^{\text{aux}} {}^{\text{R}}\mathbf{A}_{\text{FC}}]\}). \quad (48)$$

Thus, it is sufficient to include internal and rotational coordinates in the auxiliary set of coordinates. By default we apply auxiliary coordinates and transformation matrices that do not consider the translational contributions

The general transformation algorithm using the respective transformation matrices can be found in Section IIIC for PESs represented in polynomials of the coordinates. We note that whenever we transform polynomial PES representations with higher polynomial order than mode combination level, the transformation formally leads to higher-order mode combinations that are not included in the original PES representation⁵⁸. Similar arguments can also be applied the other way around: higher-order mode combinations in auxiliary coordinates could upon transformation contribute to lower-order mode combinations in the common FALCON coordinates. All together, this means that the transformation of the PESs will not give any additional error in the limit of fully expanded fragment PESs. However, this is a limit we generally have to avoid for efficiency. Since higher polynomial orders are well known to be important for accuracy, we will use these representations of the PESs for the transformation and accept the inconsistency in the treatment. For the expected typical calculation there will, thus, be an additional error source from limiting the mode-coupling expansion when employing auxiliary coordinates.

III. ALGORITHM AND IMPLEMENTATION

We report here on a pilot implementation of the double incremental schemes into MidasCpp (Molecular interactions, dynamics and simulation Chemistry program package in C++)⁵⁹. In many aspects, we have made recourse to the most fundamental rather than most elaborate schemes. For example, our incremental scheme for the evaluation of the electronic energy does not contain any multi-layer approach and we rely on simple hydrogen capping in the fragmentation. We furthermore generate the PESs on a simple static grid rather than to use adaptive schemes to choose the most important grid points and also the applied transformation algorithm is approximate as previously described. Nevertheless, the implementation provides a proof-of-principle and gives a first insight on the perspective of these schemes for large-scale PES generations. Possible beneficial extensions of this scheme for actual applications are discussed in the outlook in Section VI.

A. Partitioning and preparation

The fragments are chosen on input to the FALCON scheme. It is, however possible to obtain this input in a so-called FALCON preparation run, as also used in Ref. 54. The FALCON algorithm is then used to set up the rectilinear, semi-local FALCON coordinates. At its end, the FCR is generated, defined by a maximal fragment combination level. We furthermore by default only include those fragment combinations that are covalently bound or direct neighbors and thereby introduce spatially restricted FCRs. Then, we choose for each fragment combination those coordinates that are non-rigid within this mode combination and add hydrogen atoms in case of dangling bonds to allow sensible electronic structure calculations on each fragment combination. Finally, we generate the auxiliary inter-connecting modes and also these are appropriately capped. These steps are described in the following.

1. *Capping of dangling bonds*

The capping of dangling bonds is achieved by replacing the next atom in line with a hydrogen atom and shortening the bond lengths to average bond lengths found in organic compounds from Ref. 60. The chosen bond lengths can be found in Table S-I in the supplementary material⁵⁷. This treatment does, of course, not allow for cutting of other than single bonds. The contribution of each capping atom to the displacement vectors is set to the same value as that of the atom it is bound to. By this, we avoid artifacts in the local PESs due to bond stretches to the capping atoms. These additional contributions to the displacement vectors, of course, lift the orthonormality of the basis (if given before). The underlying assumption is that the contribution of the capping atoms to the energy of the displaced structure is similar to that to the reference structure. This means we assume that its contribution to the energy difference is small, so that the PES for the respective fragment combination can be approximated by that of the capped system.

2. *Setting up auxiliary coordinates*

For setting up the auxiliary coordinates for a particular fragment combination, we first identify the space that needs to be spanned by the auxiliary coordinates. This space is composed of the inter-connecting modes within the corresponding fragment combination that do

not correspond to common FALCON modes. For example, setting up the auxiliary coordinates for the fragment combination $\{A, B, C\}$ in the left frame in Figure 1, we see that the common FALCON inter-connecting modes of type $(A) \leftrightarrow (B)$ only have contributions from atoms in the fragment combination $\{A, B\}$ and are therefore fully contained in the fragment combination $\{A, B, C\}$. This means that we do not need to set up auxiliary coordinates describing the relative motion of $(A) \leftrightarrow (B)$. The remaining inter-connecting modes, that affect $\{A, B, C\}$, are of type $(C) \leftrightarrow (D)$, $(AB) \leftrightarrow (CD)$, and $(ABCD) \leftrightarrow (EFGH)$. Accordingly, they are not fully contained in $\{A, B, C\}$ and we need to set up the relative motions $(AB) \leftrightarrow (C)$ in the auxiliary coordinates. For this we first initialize the local translational and rotational coordinates of the relevant fragment combinations. In the above case, these are (AB) and (C) . These initial coordinates are then modified in a FALCON procedure (described in Ref. 54) using the same specifications as in the generation of the common FALCON coordinates (only on a fragment combination, that is $\{A, B, C\}$ in the present example, rather than the complete system). At the end of the FALCON procedure, reduced Hessians for sets of inter-connecting modes are diagonalized.

The Hessian, which is used during this FALCON procedure to generate auxiliary coordinates for a fragment combination, is that of the full system. This avoids potential instabilities in the procedure due to our starting structure representing only an equilibrium structure for the entire system, but not for its fragments and fragment combinations. The corresponding reduced Hessian can be calculated for the reduced space of the coordinates of interest. In principle, the same double incremental treatment described here for PESs can also be applied in (semi-numerical) Hessian calculations. Making use of such an algorithm to generate the auxiliary coordinates will make this generation of the auxiliary coordinates scale linearly with system size. This possibility is, however, not exploited in the present work.

The resulting coordinates are then used as auxiliary coordinates for the inter-connecting vibrational modes. Finally, the rotational degrees of freedom of the full fragment combination ($\{A, B, C\}$ in the present example) are added to the auxiliary coordinates. Also these diagonalize the corresponding reduced Hessian of the entire system.

B. Generation of the potential energy surface

First, the PESs of all fragment combinations are calculated separately. This is by default done in the auxiliary basis, if this reduces the number of modes to be considered. When employing auxiliary coordinates, we have to set their boundaries, which define the vibrational space covered in the PES generation, in a way that fits the vibrational space covered by the common FALCON coordinates. In practice, we set the boundaries of auxiliary coordinates ($b_{\text{FC},i}^{\text{aux}}$) from the boundaries of the common inter-connecting modes (b_j^{F}) as,

$$b_{\text{FC},i}^{\text{aux}} = \sum_{j \in \{j_{\text{max},i}\}} |(\mathbf{A}_{\text{FC}})_{ij}| b_j^{\text{F}}, \quad (49)$$

where $\{j_{\text{max},i}\}$ contains those j s that give the n largest contributions $|(\mathbf{A}_{\text{FC}})_{ij}| b_j^{\text{F}}$ for the auxiliary coordinate i , where \mathbf{A}_{FC} is the transformation matrix. n is the smaller of the maximum mode combination level and the number of common FALCON modes to be expressed in the auxiliary coordinates. Subsequent to the construction of the fragment combination's PES representation in auxiliary coordinates, it is transformed to the representation in common FALCON coordinates, see Section III C for details. After having obtained all fragment combinations' PES representations in common FALCON coordinates, the contributions of every fragment combination to the overall PES are calculated according to Eq. (12) and finally all contributions are summed up using Eq. (11).

C. Coordinate transformation of polynomial potential energy surfaces

The transformation matrix is set up as described in Section II J. In Appendix C we describe our algorithm for transforming a PES in polynomial n -mode representation in auxiliary coordinates to a polynomial n -mode representation in common coordinates. In the usual case where the order of the polynomial expansion is larger than the mode combination level, this transformation procedure can generate higher-order mode combinations that have not been included in the original representation of the PES⁵⁸. For sake of computational efficiency, we do not generate these terms.

The applied transformation algorithm scales, in case of a PES representation with up to two-mode couplings, cubic with the number of auxiliary coordinates and quadratic with the number of common FALCON coordinates that are described by auxiliary coordinates

in the original PES representation. The algorithm employed here is restricted to PES representations of polynomial form. Other algorithms are possible, which allow the usage of other analytic representations of the PES. It is for example possible to estimate the energy corresponding to displacements along the common FALCON coordinates from the PES representation in auxiliary coordinates by interpolation and then fit a new representation of the PES to these points. Since, the exact points are typically not explicitly contained in the auxiliary PES representation, interpolation techniques using derivative information (as, e.g., used in Refs. 23 and 61) will improve the description further in this case. These possibilities are, however, not a subject of the work presented here.

IV. SETUP AND COMPUTATIONAL DETAILS

The generation of the common and auxiliary FALCON coordinates and the PES construction require the accessibility of electronic energies and gradients. Most electronic structure calculations are performed with the ORCA program system⁶² employing the *Hartree–Fock with three corrections* (HF-3c) method⁶³. We performed additional calculations using density functional methods, more precisely the Becke–Perdew functional (BP86)^{64,65} and a split valence polarization (SVP) double- ζ basis set⁶⁶ using the TURBOMOLE^{67–69} program package. In these calculations, we applied the resolution-of-the-identity approximation (RI) with the corresponding auxiliary basis set⁷⁰.

The initial structures of the oligo-phenyl test systems are obtained with Avogadro⁷¹. The structures have then been optimized using the respective electronic structure method. After defining the initial groups for the FALCON procedure to be the single phenyl rings, we determine the common FALCON coordinates applying the following settings, see Ref. 54 for details. We apply distance-based coupling estimates and a degeneracy threshold of $1 \cdot 10^{-5}$ bohr⁻¹. Relaxation is included for each initial group, i.e., each fragment, meaning that the reduced Hessian for the coordinates for each group/fragment is diagonalized separately. We furthermore relax the inter-connecting coordinates separately for each type of coordinates. In this relaxation setup, only coordinates with contributions from the same fragments are mixed and therefore the overall local character (as exemplified in Figure 1) is maintained.

The fragments for the incremental PES generations are the same as the initial groups

used in the FALCON procedure. The applied FCRs only contain direct neighbors, with any atom pair closer than 3.5 bohr. We furthermore employ effective FCRs, omitting all fragment combinations with overall zero contribution.

We have applied the existing functionality in MidasCpp⁵⁹ for grid-based n -mode expansions^{56,72} for the generation of the PESs of the fragments and fragment combinations. Therefore, we have applied equidistant grids with boundaries chosen to be the classical turning points corresponding to a harmonic oscillator with quantum number of $v = 10$. These turning points are calculated using the quasi-harmonic FALCON frequencies of the respective modes. For the one-mode grids, we have applied 16 grid points per dimension, whereas we have used eight grid points per dimension, meaning 64 grid points for each two-mode cut. All PESs are approximated by an n -mode expansion up to two-mode couplings with the cut functions fitted to polynomials. The maximum polynomial order of the PES fits is in all cases set to twelve.

In the following, we apply the short-hand notations for the differently obtained PES representations: We denote full n -mode expansion by “ nM ” (only 2M in the present work) and use “DIF- $fFnM$ ” and “DIFACT- $fFnM$ ” for DIF and DIFACT expansion, respectively, up to f th order in fragment combination and n th order in mode combination. As mentioned above, the FCR is by default spatially restricted, i.e., we consider only neighbor couplings.

For the validation of the PES we have performed state-specific vibrational self-consistent field (VSCF) calculations for the one-mode excited states. These calculations were performed with the VSCF implementation²⁵ in MidasCpp⁵⁹, using a b-spline basis⁷³ of the order ten with a b-spline density of 0.8 in the region explicitly covered in the PES generation, i.e., up to the classical turning point with the harmonic quantum number $v = 10$. Again these classical turning points were calculated from the quasi-harmonic frequency of the respective FALCON modes.

V. NUMERICAL EXAMPLES FOR OLIGO-PHENYLS

For first numerical examples for the double incremental scheme, we have chosen chain-like tetra- and hexa-phenyl molecules. We employ structures with alternating dihedral angles between the phenyl entities of about ± 30 degrees (see Figure 3). These structures represent stationary points for the chosen method to calculate the electronic energy. We treat every

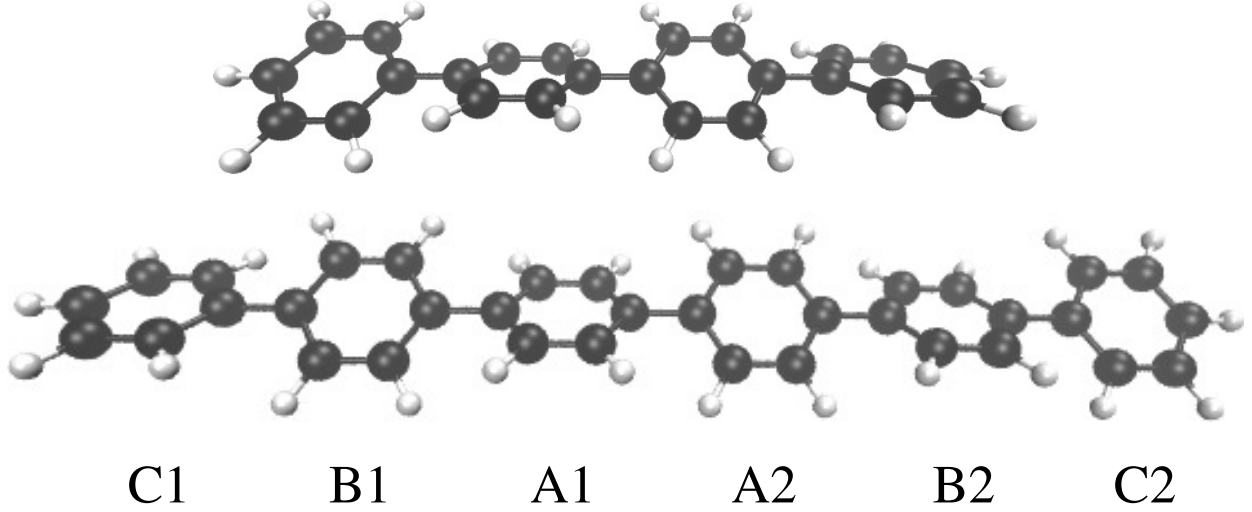


FIG. 3. Applied structure for tetra- and hexa-phenyl (HF-3c).

phenyl entity as an individual fragment and initial group of the FALCON algorithm. We apply inter-connecting FALCON coordinates for the tetra-phenyl of kind $[B1]-(A1) \leftrightarrow (A2)-[B2]$ and $(B1) \leftrightarrow (A1-A2) \leftrightarrow (B2)$ and that for the hexa-phenyl are obtained as $[C1-B1]-(A1) \leftrightarrow (A2)-[B2-C2]$, $[C1]-(B1) \leftrightarrow (A1-A2) \leftrightarrow (B2)-[C2]$, and $(C1) \leftrightarrow (B1-A1-A2-B2) \leftrightarrow (C2)$. The naming of the fragments is according to Figure 3, and the nomenclature for the classification of the different types of modes that of section II F.

The errors in PES representations due to different approximations are assessed by the difference in the state-specific VSCF energy for the fundamentally excited states from the VSCF zero-point energy.

A. Effective fragment combination ranges and computational cost

In the FCR, we only account for direct neighbors. The effective FCRs, i.e., those without zero contribution according to Eq. (15) are summarized in Table I. In these chain-like examples with spatially restricted FCR, a significant number of lower-order fragment combinations have zero contribution to the overall expansion, as shown more generally in Appendix B. This also affects the required number of SPCs and accumulated cost of all SPCs. These computational figures are shown in Figure 4 for different schemes to generate the PES. As expected from the discussion in Section II H, the number of required SPCs increases significantly with the fragment combination level. It is generally larger for the DIF model than for

TABLE I. Spatially restricted effective FCRs considering only neighbor couplings for tetra- and hexa-phenyl, with a maximal fragment combination level (f) from two to four.

f Effective FCR
Tetra-phenyl
2 $\{\{A1\}, \{A2\}, \{B1, A1\}, \{A1, A2\}, \{A2, B2\}\}$
3 $\{\{A1, A2\}, \{B1, A1, A2\}, \{A1, A2, B2\}\}$
4 $\{\{B1, A1, A2, B2\}\}$
Hexa-phenyl
2 $\{\{B1\}, \{A1\}, \{A2\}, \{B2\}, \{C1, B1\}, \{B1, A1\}, \{A1, A2\}, \{A2, B2\}, \{B2, C2\}\}$
3 $\{\{B1, A1\}, \{A1, A2\}, \{A2, B2\}, \{C1, B1, A1\}, \{B1, A1, A2\}, \{A1, A2, B2\}, \{A2, B2, C2\}\}$
4 $\{\{B1, A1, A2\}, \{A1, A2, B2\}, \{C1, B1, A1, A2\}, \{B1, A1, A2, B2\}, \{A1, A2, B2, C2\}\}$

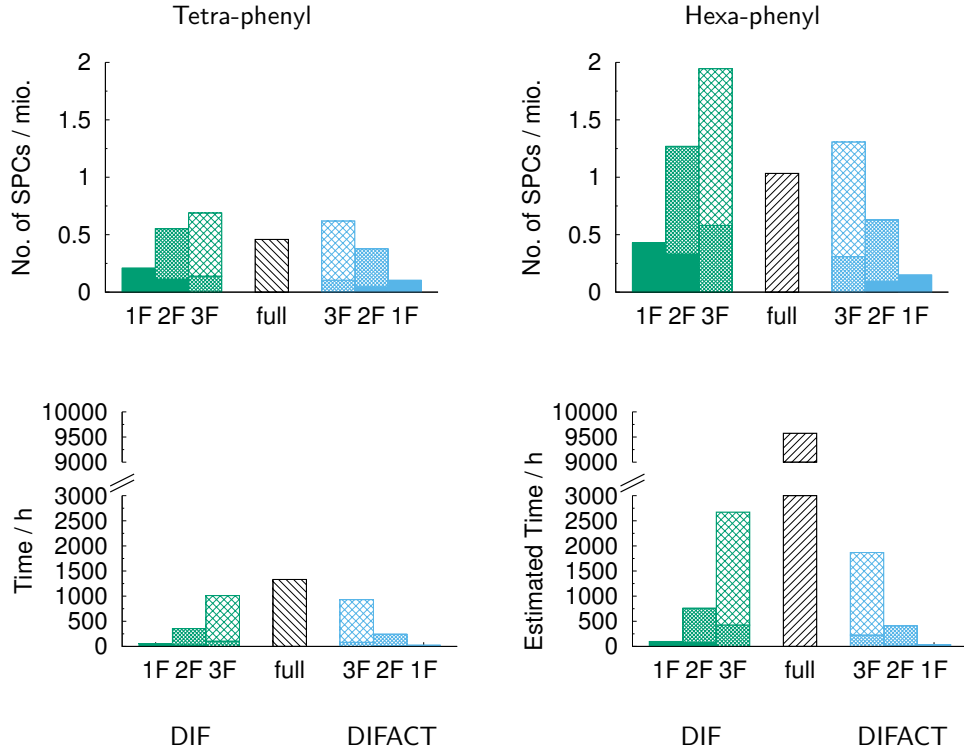


FIG. 4. Number of required SPCs and accumulated wall time of SPCs (HF-3c, on Intel Ivy Bridge 20 cores @ 2.8 GHz / 128GB; Dell C6220) for different spatially restricted effective FCRs for tetra-phenyl and hexa-phenyl in full vibrational space. The different patterns indicate, whether the contributions stem from a one-fragment potential (filled), a two-fragment potential (fine criss-cross), a three-fragment potential (coarse criss-cross), or the potential for the full four- or six-fragment system (feint ruled). The wall times for hexa-phenyl are estimated (see main text).

the DIFACT approach with the same fragment combination level. The difference between these two models is, as expected, significantly larger for the hexa-phenyl than for the tetra-phenyl example. The number of required SPCs in double incremental PES constructions is in many cases larger than the required number of SPCs for generating the full 2M PES.

To give an impression on how the computational cost differs between the different approaches, we have timed the individual SPCs on nodes of the type Intel Ivy Bridge 20 cores @ 2.8 GHz / 128GB; Dell C6220. The times plotted for the hexa-phenyl are estimated from the number of SPCs and averaged wall time for several SPCs of displaced structures of the respective molecules. This means, that the timings reported in the bottom part of Figure 4 should be considered ballpark figures rather than rigorous assessments. They can still give valuable insight into the computational gains in the double incremental scheme in actual computations. First of all, we see that all shown double incremental expansions require less computational resources than the 2M reference calculations. As expected, this difference is more pronounced for the hexa-phenyl than for the tetra-phenyl example. The accumulated computational cost for SPCs is typically the bottleneck in PES generations. The double incremental approaches, therefore, offer a large gain in the feasibility of PES generations for large molecular systems. The relative gain is expected to be more beneficial the larger the investigated system is (see also Section II H).

B. Error assessment

In this section, we assess how well we can reproduce the conventional, non-incremented calculation with the different approximate double incremental schemes in FALCON coordinates. The errors to be assessed are (i) the error introduced by the truncated double incremental expansion, including the capping of dangling bonds and (ii) a transformation error, in case auxiliary modes are employed.

1. *Error assessment for inter-connecting modes*

For the first error assessment, we restrict the discussion to the vibrational subspace of inter-connecting modes in the tetra- and hexa-phenyl examples. For these modes, the errors introduced by the double incremental expansion and usage of auxiliary coordinates are

expected to be particularly large, since (i) these modes include often more fragments than covered in the largest fragment combination level and (ii) the transformation procedure is only applied to these modes.

Table II shows the results obtained considering only these inter-connecting modes for tetra-phenyl and employing HF-3c for the SPCs. First we assess the capping and fragmentation error (i): We see that the root mean square deviation (RMSD) of the fundamental VSCF excitation energies obtained with a DIF-2F2M potential compared to the 2M reference amounts to only 0.54 cm^{-1} and the maximum absolute deviation is 1.31 cm^{-1} . The results for the DIF-3F2M potential and the reference potential are in almost perfect agreement with a maximum absolute deviation in fundamental excitation energies of 0.05 cm^{-1} . The absolute deviations of the VSCF zero-point energy obtained with DIF-2F2M and DIF-3F2M, respectively, are only slightly larger than maximum absolute deviation of the excitation energies. It amounts to 1.96 cm^{-1} for DIF-2F2M and to 0.05 cm^{-1} for DIF-3F2M. These findings suggest very good and quick convergence of the fragmentation error with increasing fragment combination level.

Next, we turn to the transformation error (ii), i.e., the deviation between the fundamental excitation energies obtained with a PES generated in auxiliary coordinates from those for the corresponding PES representation without this approximation. We see that the RMSD caused by this error lies around 0.7 cm^{-1} both for DIFACT-2F2M and DIFACT-3F2M and the maximum absolute deviation is at 1.81 cm^{-1} for DIFACT-2F2M and at 2.11 cm^{-1} for DIFACT-3F2M.

Very similar, though slightly larger errors caused by the fragmentation and transformation are obtained for the larger hexa-phenyl molecule (see Table III and Table S-II in the supplementary material⁵⁷ for full details), where we again consider only inter-connecting modes. All in all, we see also here quick convergence of the double incremental scheme with almost perfect agreement already for the DIF-3F2M potential with a maximum absolute deviation in the fundamental excitation energies of 0.13 cm^{-1} . Concerning the transformation error, we obtain RMSDs between 0.66 and 0.88 cm^{-1} for fragment combinations levels of two to four. These errors are, hence, rather close to those for the tetra-phenyl example with only inter-connecting modes (see Table II).

TABLE II. Zero-point energies (ZPEs) and differences of the fundamental energies from the ZPE (ΔE) obtained from state-specific VSCF calculations for tetra-phenyl for different PES representations (HF-3c) and considering only inter-connecting modes. ω are the corresponding quasi-harmonic FALCON frequencies. All energies are given in cm^{-1} . The type label IC1 refers to $[\text{B1}]\text{-(A1)}\leftrightarrow(\text{A2})\text{-[B2]}$ and IC2 to $(\text{B1})\leftrightarrow(\text{A1-A2})\leftrightarrow(\text{B2})$ type of FALCON coordinate.

f	Type	ω	2M	DIF- f F2M		DIFACT- f F2M	
				2	3	2	3
ZPE			2818.97	2817.01	2818.92	2817.70	2819.15
Fundamental excitation energies							
1	IC2	45.95	56.96	57.03	56.96	58.01	56.89
2	IC2	59.23	87.30	87.49	87.32	85.68	89.40
3	IC2	74.54	82.12	82.27	82.12	82.57	82.49
4	IC2	83.84	102.26	102.63	102.26	102.63	100.15
5	IC2	87.92	97.71	97.67	97.70	98.73	97.84
6	IC1	102.51	114.85	115.43	114.87	116.89	115.72
7	IC2	132.33	138.46	137.96	138.45	138.16	138.65
8	IC1	326.79	329.03	329.30	329.03	329.81	329.29
9	IC2	349.13	352.29	352.53	352.28	352.54	352.03
10	IC1	370.57	371.80	371.15	371.77	370.95	371.97
11	IC2	376.65	374.75	374.55	374.75	374.53	374.76
12	IC2	401.39	403.34	404.43	403.39	404.70	403.15
13	IC2	428.26	429.10	428.76	429.07	429.02	429.02
14	IC2	499.72	499.83	499.48	499.83	499.75	499.67
15	IC1	503.33	504.05	503.27	504.07	503.29	504.23
16	IC2	534.47	531.13	530.82	531.10	530.72	531.09
17	IC1	549.26	549.31	549.10	549.30	548.91	549.37
18	IC1	658.01	654.06	652.75	654.01	652.71	654.02
Δ ZPE to							
2M				-1.96	-0.05	-1.27	0.19
DIF- f F2M						0.69	0.24
Root mean square deviation of ΔE to							
2M				0.54	0.02	0.93	0.75
DIF- f F2M						0.67	0.74
Maximum absolute deviation of ΔE from							
2M				1.31	0.05	2.04	2.11
DIF- f F2M						1.81	2.11

TABLE III. Zero-point energies (ZPEs) as well as root mean square deviation and maximum absolute deviation of the difference in VSCF state energies for singly excited states from the ZPE (ΔE) for hexa-phenyl for different PES representations (HF-3c) and considering only inter-connecting modes. All energies are given in cm^{-1} .

	2M	DIF- <i>f</i> F2M			DIFACT- <i>f</i> F2M		
<i>f</i>		2	3	4	2	3	4
ZPE	4797.68	4794.19	4797.55	4797.67	4796.02	4798.16	4797.62
Δ ZPE to							
2M		-3.49	-0.12	-0.00	-1.66	0.48	-0.06
DIF- <i>f</i> F2M					1.83	0.60	-0.06
Root mean square deviation of ΔE to							
2M		0.80	0.04	0.00	1.25	0.88	0.66
DIF- <i>f</i> F2M					0.75	0.88	0.66
Maximum absolute deviation of ΔE from							
2M		2.11	0.13	0.01	3.55	2.96	2.82
DIF- <i>f</i> F2M					2.13	2.92	2.81

2. Error assessment in full space

Considering all 120 vibrational degrees of freedom of the tetra-phenyl example, we obtain very similar fragmentation errors compared to the above case with only inter-connecting modes (compare Tables II and IV, see also S-III in the supplementary material⁵⁷ for all 120 fundamental excitation energies for tetra-phenyl using PESs obtained with HF-3c). The deviations found for the zero-point energy are again larger than those for the fundamental excitation energies. Both decrease systematically when increasing the maximal order of fragment combination in the DIF scheme. The maximum absolute deviation of the energy differences obtained with DIF-2F2M from the 2M reference corresponds to an intra-fragment mode and amounts to 2.35 cm^{-1} . The maximum absolute deviation for the inter-connecting modes is 1.37 cm^{-1} in this case. For DIF-3F2M, all fundamental excitation energies agree within $\pm 0.1 \text{ cm}^{-1}$. As expected, the transformation error is generally larger for the directly affected inter-connecting modes than for the intra-fragment modes. The latter are only indirectly affected by the transformation via the coupling elements to the inter-connecting modes. In the present case, the transformation error decreases for intra-fragment modes and increases for inter-connecting modes when going from DIFACT-2F2M to DIFACT-3F2M.

TABLE IV. Zero-point energies (ZPEs) as well as root mean square deviation (RMSD) and maximum absolute deviation (MAX) of the difference in VSCF state energies for singly excited states from the ZPE (ΔE) for tetra-phenyl for different PES representations (HF-3c). The deviations for inter-connecting (IC) modes and intra-fragment (INTRA) modes are shown separately and combined (All). All energies are given in cm^{-1} .

	2M	DIF- <i>f</i> F2M		DIFACT- <i>f</i> F2M	
<i>f</i>		2	3	2	3
ZPE	86266.64	86263.37	86266.54	86263.54	86267.29
Δ ZPE to 2M					
		-3.27	-0.10	-3.10	0.65
Δ ZPE to DIF- <i>f</i> F2M					
				0.17	0.76
RMSD of ΔE to 2M					
All		0.56	0.02	0.76	0.59
IC		0.51	0.02	1.10	1.51
INTRA		0.57	0.02	0.69	0.12
RMSD of ΔE to DIF- <i>f</i> F2M					
All				0.44	0.59
IC				0.82	1.51
INTRA				0.32	0.12
MAX of ΔE from 2M					
IC		1.37	0.05	2.12	4.79
INTRA		2.35	0.09	2.65	0.29
MAX of ΔE from DIF- <i>f</i> F2M					
IC				1.79	4.79
INTRA				0.57	0.28

Figure 5 shows the error distributions for fragmentation, transformation, and overall error for the example of PES representations up to second order in fragment combination. The error distributions are well centered on zero. The overall error has similar contributions from the fragmentation and transformation error in this case, as has already been suggested by the similar RMSDs. This supports that the transformation at the 2F level does not lead to an inherent larger error for the full procedure. An analogous analysis for the 3F case (not shown) yields an almost perfect agreement of the DIF-3F2M results with the 2M reference results and also the transformation error for most fundamental excitation energies, lies close

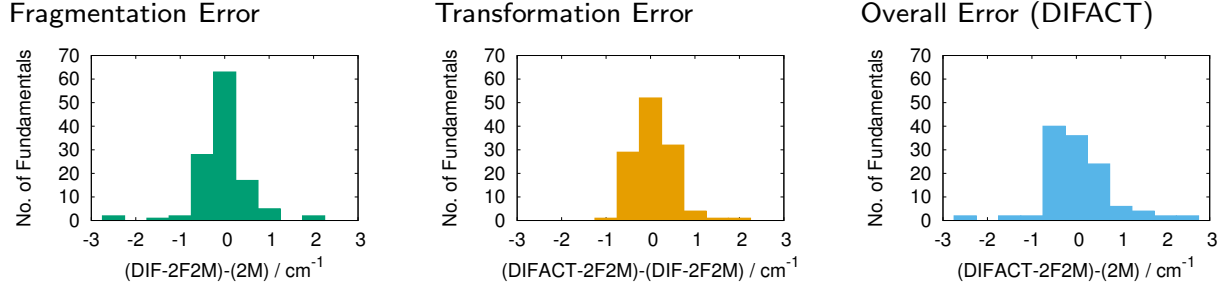


FIG. 5. Error distributions for the VSCF fundamental excitation energies for DIF-2F2M and DIFACT-2F2M representations of the PES (HF-3c) for tetra-phenyl in full vibrational space with respect to each other (middle graph) and the reference results for the 2M PES representation (left and right graph, respectively).

to zero: Only three fundamental VSCF excitation energies obtained with a DIFACT-3F2M representation of the PES deviate by more than 0.6 cm^{-1} from the DIF-3F2M results. Those have an absolute deviation of 3.71 cm^{-1} , 1.88 cm^{-1} , and 4.79 cm^{-1} and correspond to vibrations of the type $(\text{B1}) \leftrightarrow (\text{A1-A2}) \leftrightarrow (\text{B2})$ with quasi-harmonic frequencies clearly below 100 cm^{-1} . This means that the comparatively large maximum absolute transformation error of 4.79 cm^{-1} for DIFACT-3F2M is likely an outlier. This suggests, that only these three common FALCON coordinates are not very well resembled in the auxiliary coordinates of the fragment combinations in the present case, while the PES contributions of all others are surprisingly well maintained in the potential transformations.

Next, we turn to the hexa-phenyl example, whose results are summarized in Table V. The full data set of calculated excitation energies for hexa-phenyl can be found in Table S-IV in the supplementary material⁵⁷. Again the largest deviation of -6.94 cm^{-1} is obtained for the zero-point energy when including fragment combinations up to second order. The deviations in the fundamental excitation energies obtained in the double incremental schemes from the 2M reference results for hexa-phenyl are of similar size as those for the tetra-phenyl discussed above (see Table IV), though slightly larger. The DIF-3F2M results agree again very well with the reference results, with a maximum deviation in fundamental excitation energies of 0.22 cm^{-1} (see Table V). Again, the majority of the fundamental excitation energies have transformation errors below 1 cm^{-1} and only few fundamental excitation energies, which all correspond to low-lying inter-connecting modes, exhibit a considerably larger transformation

TABLE V. Zero-point energies (ZPEs) as well as root mean square deviation (RMSD) and maximum absolute deviation (MAX) of the difference in VSCF state energies for singly excited states from the ZPE (ΔE) for hexa-phenyl for different PES representations (HF-3c). The deviations for inter-connecting (IC) modes and intra-fragment (INTRA) modes are shown separately and combined (All).

f	2M	DIF- f F2M		DIFACT- f F2M	
		2	3	2	3
ZPE	127147.62	127140.68	127147.33	127142.81	127147.80
Δ ZPE to 2M		-6.94	-0.29	-4.81	0.17
Δ ZPE to DIF- f F2M				2.13	0.47
RMSD of ΔE to 2M					
All		0.75	0.04	0.97	0.65
IC		0.77	0.04	1.46	1.54
INTRA		0.75	0.04	0.84	0.17
RMSD of ΔE from DIF- f F2M					
All				0.49	0.65
IC				0.99	1.54
INTRA				0.30	0.16
MAX of ΔE from 2M					
IC		2.17	0.13	4.34	4.83
INTRA		3.13	0.22	3.12	0.45
MAX of ΔE from DIF- f F2M					
IC				3.14	4.83
INTRA				0.65	0.42

error. The maximum absolute deviation due to the potential transformation amounts to 4.83 cm^{-1} in this case.

We have, thus, found in this first numerical study that the double incremental scheme can indeed yield very good agreement with the supermolecular calculations. Moreover, the convergence with respect to fragment combination order is rather quick and already the results for the three-fragment combinations are almost perfectly in line with the reference values. The results illustrate that the semi-local FALCON coordinates provide a good set of

coordinates that allows distance cutoff in the PES construction with fast convergence. The controlled and well-defined locality of the FALCON coordinates is essential for the efficiency of DIF procedure.

We see that the transformation error is somewhat similar in all cases. It is generally lower for intra-fragment modes than for inter-connecting modes. There is no clear tendency for the transformation error, when going to higher-order fragment combinations. However, this is also what should be expected. What is required to reduce the transformation errors is a better transformation such as a better algorithm and/or a higher level in mode coupling. Thus, a 3M expansion and transformation would allow the transformed potential to capture better the correct features of the true potential. However, already for the 2M case, we observe a transformation error of less than 5 cm^{-1} in all fundamental energies, and the majority of cases have errors less than 1 cm^{-1} . These errors are modest, both in view of the approximations involved in the present transformation algorithm applied in the DIFACT scheme as described earlier, but also in the sense of being modest compared to other sources of errors, including the 2M restriction, errors from the electronic structure methodology applied, etc. The good performance in the PES transformation can, at least partly, be rationalized by the very similar fashion in which the common and auxiliary modes are set up. In this way, we generate auxiliary coordinates that in many cases resemble the common modes to a large extent, and thereby only modest coordinate rotations take place in the potential transformation. Modest coordinate rotations are expected to lead to higher accuracy in the current potential transformation algorithm. Other algorithms may be less sensitive with respect to this aspect. Recalling that this additional transformation enables linear scaling of the accumulated computational cost of the SPCs in the PES generation, this error seems legitimate, especially when interested in excitations dominated by intra-fragment modes. Still if one needs more accurate potentials, the DIF scheme also offers a significant reduction of the computational cost compared to the complete calculation and has the advantage of clear and rigorous limits towards the full expansion.

3. Change of electronic structure method

In Sections V B 1 – V B 2, we have seen that the error that we introduce by fragmentation in the double incremental scheme lies, even for the DIFACT-2F2M approximation, below

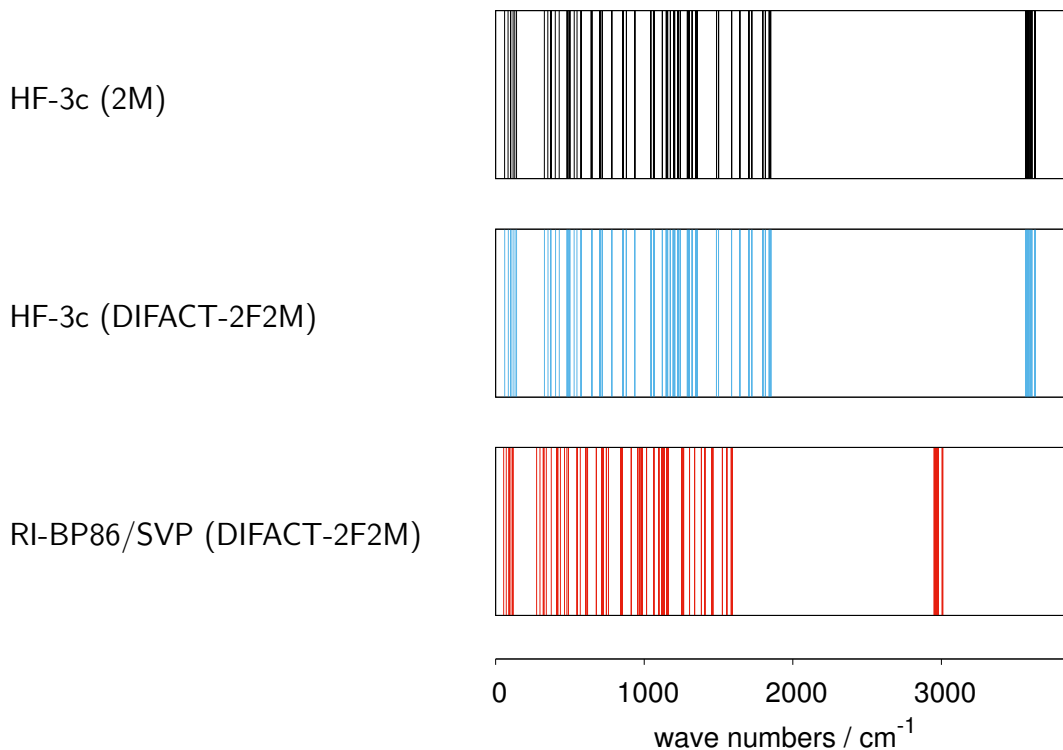


FIG. 6. Stick spectra of VSCF fundamental excitation energies in the applied FALCON coordinates for HF-3c and RI-BP86/SVP in selected representations of the PES, i.e., 2M and DIFACT-2F2M.

5 cm^{-1} in all cases and is thereby below the error made through other approximations, such as the electronic structure method. This suggests, that it is likely beneficial to use a more reliable (and thereby more expensive) electronic structure method. Here, we illustrate that the DIFACT scheme indeed makes it feasible to use more accurate electronic structure methods in the PES construction with more than hundred degrees of freedom. We compare the DIFACT-2F2M potentials for tetra-phenyl obtained with the HF-3c to that from RI-BP86/SVP SPCs. The respective stick spectra are together with a reference 2M spectrum for HF-3c shown in Figure 6. The respective data can be found in Tables S-III and S-V in the supplementary material⁵⁷. It is striking that the differences between the different PES generation schemes using HF-3c is hardly visible, whereas the difference due to the different electronic structure method is rather significant. It should also be mentioned that the frequently applied frequency scaling for the two methods compared here is rather different: In Ref. 63, a frequency scaling factor of 0.86 is applied for HF-3c frequencies,

whereas in a recent study fitting different frequency scaling factors⁷⁴, a frequency scaling factor of 1.03 for BP86/def2-SVP is advertised. Note that def2-SVP is equivalent to SVP for carbon and hydrogen atoms, i.e., the only atoms contained in the oligo-phenyl examples. This means we cannot really expect the VSCF results for PESs obtained with these two electronic structure methods to be similar. Nevertheless, the strong effect of the choice of the electronic structure method compared to the errors introduced by the double incremental expansion is remarkable. With the DIF and DIFACT schemes methods one has new methods for finding good compromises between accuracy and efficiency. Thus, a double incremental PES expansion with a more expensive electronic structure method may be a more promising route to efficient generation of reliable PESs for vibrational structure calculations for larger systems, than to compromise the potential by using a lower-level electronic structure method, or avoiding higher-level mode couplings.

VI. SUMMARY, CONCLUSIONS AND OUTLOOK

We have presented a double incremental expansion for the generation of PES for vibrational structure calculations combining the n -mode expansion of the PES with the many-body expansion of the underlying evaluations of the electronic energies. We have shown that this type of double incremental expansion of the PES is particularly beneficial when combined with semi-local coordinates to span the vibrational space. This leads to a significant reduction of the number of required SPCs. For this reason, we use the recently introduced rectilinear and semi-local FALCON coordinates⁵⁴ in combination with the double incremental expansion leading to the double incremental scheme in FALCON coordinates (DIF). Even in these coordinates, the number of required modes to be considered for the PES generation for fragment combinations often exceeds the number of vibrational degrees of freedom in the respective fragment combination. This means, that the PESs of the fragment combinations are spanned in a redundant basis and the explicit calculation of grid points in this redundant basis leads to similar overall scaling behavior as in delocalized modes. This scaling behavior of the DIF scheme is caused by a limited number of modes. These are so-called inter-connecting modes, i.e., those modes that describe the relative motion of the fragments. In an alternative scheme, we can express these motions by auxiliary modes. In this scheme, the generation of the PES for every fragment combination only considers the number of

vibrational degrees of freedom in the respective fragment combination plus three modes corresponding to rotations. In this way, the number of required SPCs per fragment combination is independent of the total number of fragments and an overall low scaling behavior can be achieved. This is, however, achieved at the cost of introducing an additional error in the subsequent transformation of the fragment combinations’ PESs to the common coordinates. We stress that the scaling analysis does not only consider the number of required SPCs, but also takes the cost per SPC into account. No assumptions of a particular scaling of the individual SPCs with system size has been made. Our approach differs from previous low scaling approaches for the generation of PESs^{25,28}, where “linear scaling” refers to the number of required SPCs but not to the accumulated computational cost of these SPCs.

Numerical examples on tetra- and hexa-phenyls show quick convergence of the VSCF fundamental excitation energies for PESs constructed with increasing level in fragment combination. Indeed the DIF-3F2M results resemble the 2M results almost perfectly in the tested cases. Using auxiliary coordinates, i.e., the corresponding DIFACT scheme, leads to additional RMSDs of below 1 cm^{-1} with maximum absolute deviations of less than 5 cm^{-1} . The larger transformation errors are, in these examples, restricted to very few fundamental excitation energies that correspond to inter-connecting modes. The overall modest size of these errors can, at least partly, be related to relative close resemblance of the overall FALCON modes to the auxiliary modes of the same type. We have illustrated that this error lies clearly below the difference in fundamental excitation energies obtained with different electronic structure methods realistically applicable in the generation of PESs for systems containing tens of atoms or more. This suggests that it may be more relevant to choose a more accurate and thereby a computationally more demanding electronic structure method, but at low order in the double incremental expansion to yield the most beneficial balance of computational cost and accuracy. We foresee that the computationally cheapest double incremental expansion, i.e., DIFACT-2F2M may be sufficient in many cases and will enable reliable generation of PESs for wave function-based calculations of vibrational spectra, for system sizes, that are currently clearly out of reach.

The considered test systems are perfect examples for which our methods are appropriate, as chain-like systems with modest long-range electrostatic interactions and capping adequately handled by hydrogens. In addition, the level of the electronic structure description applied was rudimentary. For many other complex systems there may be a slower decay

in importance with distance between fragments as well as more advanced capping schemes may be preferred. It is an essential ingredient of the method that only changes in energy relative to a reference structure are considered. Thus, even if the caps do not cancel for a given structure, their contribution may cancel out in the contributions to the PES. This is likely to simplify the use of other capping schemes as well, as long as rigid caps are used. Nevertheless, it must be carefully considered which steps need to be included in applying the method to complex systems with advanced cappings. Likewise, the rate of decay with distance for the fragment couplings will vary between systems, and potentially also between different computational methods. For example, the accuracy of the fragment many-body expansion has been suggested to depend on the character of the applied basis set and potentially involves a basis set superposition error.⁷⁵⁻⁷⁷

In addition to the above-mentioned examples much other work is needed extending the methodology and acquiring experience before the double incremental approach can become a widely used standard procedure. The present work focuses on PESs and the extension to other property surfaces such as dipole surfaces, is needed to calculate infrared spectra. Generally, the extension to size extensive, additive properties will be straightforward. Another point of possible improvement concerns the currently applied transformation algorithm. Other more elaborate schemes, for instance using interpolation techniques may reduce the transformation error and thereby extend the applicability of the very efficient DIFACT scheme. For efficiency and increasing the black box nature, it is also important to integrate the double incremental approach with automatic choice of required grid points (as, for instance, done in the adaptive density-guided approach (ADGA)¹⁷). Multi-level^{12,20-24} and screening approaches¹¹⁻¹⁶ as well as various interpolation and extrapolation schemes^{12,23,61} can likewise be integrated with the double incremental approach. To combine these ideas in a coherent framework will be topics for future research. Finally, the methodology developed here should also be well applicable using curvilinear internal coordinates, and at least some aspects of the theory will be simpler, such as the definition of locality, and rigidity under relative rotations etc. With the nontrivial construction of the appropriate kinetic energy operator for such coordinates becoming more automatic through recent research^{78,79}, the development of a double incremental scheme in curvilinear coordinates is a challenging but not unrealistic topic of future research.

All in all, the presented double incremental scheme for the generation of PESs pushes

size limitation for PES generation with quantum chemical evaluation of the electronic energy points to significantly larger sizes compared to the conventional n -mode expansion. Combined with other schemes to gain more efficiency, this method will be an important step towards reliable PESs and thereby towards vibrational structure calculations for sizeable systems, including covalently bound molecules. We expect our approach to be a valuable complement to vibrational wave function approaches for bio-molecular vibrational spectroscopy^{26,80–82} as well as molecular vibrations in heterogeneous environments, such as molecules on metal surfaces^{83–85}.

ACKNOWLEDGMENTS

We are grateful to Ian H. Godtlielsen for continuous support in the development of the MidasCpp code as well as Emil Lund Klinting, Mads Bøttger Hansen, Diana Madsen, Niels Kristian Madsen, and Gunnar Schmitz for careful reading of this manuscript. C.K. acknowledges funding by a Feodor–Lynen research fellowship from the Alexander von Humboldt Foundation as well as a post-doctoral grant from the Carlsberg Foundation. O.C. acknowledges support from the Lundbeck Foundation, the Danish e-infrastructure Cooperation (DeiC), and the Danish Council for Independent Research through a Sapere Aude III grant (DFF – 4002-00015).

REFERENCES

- ¹J. O. Jung and R. B. Gerber, J. Chem. Phys. **105**, 10332 (1996).
- ²A. E. Roitberg, R. B. Gerber, and M. A. Ratner, J. Phys. Chem. B **101**, 1700 (1997).
- ³S. Carter, S. J. Culik, and J. M. Bowman, J. Chem. Phys. **107**, 10458 (1997).
- ⁴G. M. Chaban, J. O. Jung, and R. B. Gerber, J. Chem. Phys. **111**, 1823 (1999).
- ⁵K. Yagi, T. Taketsugu, K. Hirao, and M. S. Gordon, J. Chem. Phys. **113**, 1005 (2000).
- ⁶V. Deev and M. A. Collins, J. Chem. Phys. **122**, 154102 (2005).
- ⁷G. C. Schatz, Rev. Mod. Phys. **61**, 669 (1989).
- ⁸H. Rabitz and O. F. Aliş, J. Math. Chem. **25**, 197 (1999).
- ⁹H.-D. Meyer, WIREs: Comput. Mol. Sci. **2**, 351 (2012).

- ¹⁰M. Griebel, in *Foundations of Computational Mathematics (FoCM05), Santander*, edited by L. Pardo, A. Pinkus, E. Suli, and M. Todd (Cambridge University Press, 2006) pp. 106–161.
- ¹¹D. M. Benoit, J. Chem. Phys. **120**, 562 (2004).
- ¹²G. Rauhut, J. Chem. Phys. **121**, 9313 (2004).
- ¹³D. M. Benoit, J. Chem. Phys. **125**, 244110 (2006).
- ¹⁴L. Pele and R. B. Gerber, J. Chem. Phys. **128**, 165105 (2008).
- ¹⁵D. M. Benoit, J. Chem. Phys. **129**, 234304 (2008).
- ¹⁶P. Seidler, T. Kaga, K. Yagi, O. Christiansen, and K. Hirao, Chem. Phys. Lett. **483**, 138 (2009).
- ¹⁷M. Sparta, D. Toffoli, and O. Christiansen, Theor. Chem. Acc. **123**, 413 (2009).
- ¹⁸F. Richter, P. Carbonniere, A. Dargelos, and C. Pouchan, J. Chem. Phys. **136**, 224105 (2012).
- ¹⁹D. Strobusch and C. Scheurer, J. Chem. Phys. **140**, 074111 (2014).
- ²⁰K. Yagi, S. Hirata, and K. Hirao, Theor. Chem. Acc. **118**, 681 (2007).
- ²¹G. Rauhut and T. Hrenar, Chem. Phys. **346**, 160 (2008).
- ²²M. Sparta, I.-M. Høyvik, D. Toffoli, and O. Christiansen, J. Phys. Chem. A **113**, 8712 (2009).
- ²³M. Sparta, M. B. Hansen, E. Matito, D. Toffoli, and O. Christiansen, J. Chem. Theory Comput. **6**, 3162 (2010).
- ²⁴P. Meier, G. Bellchambers, J. Klepp, F. R. Manby, and G. Rauhut, Phys. Chem. Chem. Phys. **15**, 10233 (2013).
- ²⁵M. B. Hansen, M. Sparta, P. Seidler, D. Toffoli, and O. Christiansen, J. Chem. Theory Comput. **6**, 235 (2010).
- ²⁶P. T. Panek and C. R. Jacob, ChemPhysChem **15**, 3365 (2014).
- ²⁷X. Cheng and R. P. Steele, J. Chem. Phys. **141**, 104105 (2014).
- ²⁸F. Richter, P. Carbonniere, and C. Pouchan, Int. J. Quan. Chem. **114**, 1401 (2014).
- ²⁹P. T. Panek and C. R. Jacob, J. Chem. Phys. **144**, 164111 (2016).
- ³⁰M. A. Collins and R. P. A. Bettens, Chem. Rev. **115**, 5607 (2015).
- ³¹H. Fujimoto, N. Koga, and K. Fukui, J. Am. Chem. Soc. **103**, 7452 (1981).
- ³²K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).

- ³³D. W. Zhang and J. Z. H. Zhang, J. Chem. Phys. **119**, 3599 (2003).
- ³⁴L. Huang, L. Massa, and J. Karle, Int. J. Quant. Chem. **103**, 808 (2005).
- ³⁵W. Li, S. Li, and Y. Jiang, J. Phys. Chem. A **111**, 2193 (2007).
- ³⁶D. G. Fedorov and K. Kitaura, in *Modern Methods for Theoretical Physical Chemistry of Biopolymers*, edited by E. B. Starikov, J. P. Lewis, and S. Tanaka (Elsevier B.V., 2006) pp. 3–38.
- ³⁷D. G. Fedorov, T. Ishida, and K. Kitaura, J. Phys. Chem. A **109**, 2638 (2005).
- ³⁸Y. Mei, C. Ji, and J. Z. H. Zhang, J. Chem. Phys. **125**, 094906 (2006).
- ³⁹E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 46 (2007).
- ⁴⁰J. M. Mullin, L. B. Roskop, S. R. Pruitt, M. A. Collins, and M. S. Gordon, J. Phys. Chem. A **113**, 10040 (2009).
- ⁴¹G. J. O. Beran, J. Chem. Phys. **130**, 164115 (2009).
- ⁴²N. J. Mayhall and K. Raghavachari, J. Chem. Theory Comput. **7**, 1336 (2011).
- ⁴³J. Řezáč and D. R. Salahub, J. Chem. Theory Comput. **6**, 91 (2010).
- ⁴⁴M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).
- ⁴⁵K. Raghavachari and A. Saha, Chem. Rev. **115**, 5643 (2015).
- ⁴⁶R. M. Richard and J. M. Herbert, J. Chem. Phys. **137**, 064113 (2012).
- ⁴⁷N. J. Mayhall and K. Raghavachari, J. Chem. Theory Comput. **8**, 2669 (2012).
- ⁴⁸R. M. Richard and J. M. Herbert, J. Chem. Theory Comput. **9**, 1408 (2013).
- ⁴⁹A. J. C. Varandas and J. N. Murrell, Faraday Discuss. Chem. Soc. **62**, 92 (1977).
- ⁵⁰A. J. C. Varandas, Mol. Phys. **53**, 1303 (1984).
- ⁵¹Y. Wang and J. M. Bowman, Chem. Phys. Lett. **491**, 1 (2010).
- ⁵²Y. Wang and J. M. Bowman, J. Chem. Phys. **136**, 144113 (2012).
- ⁵³J. M. Bowman, Y. Wang, H. Liu, and J. S. Mancini, J. Phys. Chem. Lett. **6**, 366 (2015).
- ⁵⁴C. König, M. B. Hansen, I. H. Godtliebsen, and O. Christiansen, J. Chem. Phys. **144**, 074108 (2016).
- ⁵⁵S. S. Xantheas, J. Chem. Phys. **100**, 7523 (1994).
- ⁵⁶J. Kongsted and O. Christiansen, J. Chem. Phys. **125**, 124108 (2006).
- ⁵⁷See supplementary material for graphs on the scaling behavior of the estimated computational cost of the DIF and DIFACT scheme for different scalings in the electronic structure method and different maximum fragment combination level as well as a list of the bond

- lengths for capping atoms and the fundamental VSCF excitation energies for tetra- and hexa-phenyls obtained with the different representations of the potential energy surface.
- ⁵⁸K. Yagi, M. Keçeli, and S. Hirata, J. Chem. Phys. **137**, 204118 (2012).
- ⁵⁹O. Christiansen, I. H. Godtlielsen, W. Györfy, M. B. Hansen, M. B. Hansen, J. Kongsted, E. L. Klinting, C. König, S. A. Losilla, D. Madsen, N. K. Madsen, E. Matito, P. Seidler, K. Snedkov, M. Sparta, B. Thomsen, D. Toffoli, and A. Zocante, *MidasCpp (Molecular Interactions, dynamics and simulation Chemistry program package in C++)* (University of Aarhus, 2016) www.chem.au.dk/midas.
- ⁶⁰F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, and R. Taylor, J. Chem. Soc., Perkin Trans. 2, S1 (1987).
- ⁶¹E. Matito, D. Toffoli, and O. Christiansen, J. Chem. Phys. **130**, 134104 (2009).
- ⁶²F. Neese, WIREs: Comput. Mol. Sci. **2**, 73 (2012).
- ⁶³R. Sure and S. Grimme, J. Comput. Chem. **34**, 1672 (2013).
- ⁶⁴A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
- ⁶⁵J. P. Perdew, Phys. Rev. B **33**, 8822 (1986).
- ⁶⁶A. Schäfer, H. Horn, and R. Ahlrichs, J. Chem. Phys. **97**, 2571 (1992).
- ⁶⁷R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, Chem. Phys. Lett. **162**, 165 (1989).
- ⁶⁸F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, WIREs: Comput. Mol. Sci. **4**, 91 (2014).
- ⁶⁹“TURBOMOLE V5.10 2008, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH 1989–2007, TURBOMOLE GmbH, since 2007 ; available from <http://www.turbomole.com>.” (2008).
- ⁷⁰K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, Chem. Phys. Lett. **242**, 652 (1995).
- ⁷¹M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, J. Cheminform. **4**, 1 (2012).
- ⁷²D. Toffoli, J. Kongsted, and O. Christiansen, J. Chem. Phys. **127**, 204106 (2007).
- ⁷³D. Toffoli, M. Sparta, and O. Christiansen, Mol. Phys. **109**, 673 (2011).
- ⁷⁴M. K. Kesharwani, B. Brauer, and J. M. L. Martin, J. Phys. Chem. A **119**, 1701 (2015).
- ⁷⁵J. F. Ouyang, M. W. Cvitkovic, and R. P. A. Bettens, J. Chem. Theory Comput. **10**, 3699 (2014).

- ⁷⁶D. Yuan, X. Shen, W. Li, and S. Li, Phys. Chem. Chem. Phys. **18**, 16491 (2016).
- ⁷⁷K. U. Lao, K.-Y. Liu, R. M. Richard, and J. M. Herbert, J. Chem. Phys. **144**, 164105 (2016).
- ⁷⁸D. Lauvergnat and A. Nauts, J. Chem. Phys. **116**, 8560 (2002).
- ⁷⁹M. Ndong, L. Joubert-Doriol, H.-D. Meyer, A. Nauts, F. Gatti, and D. Lauvergnat, J. Chem. Phys. **136**, 034107 (2012).
- ⁸⁰A. A. Adesokan, E. Fredj, E. C. Brown, and R. B. Gerber, Mol. Phys. **103**, 1505 (2005).
- ⁸¹T. K. Roy, V. Kopysov, N. S. Nagornova, T. R. Rizzo, O. V. Boyarkin, and R. B. Gerber, ChemPhysChem **16**, 1374 (2015).
- ⁸²T. K. Roy, R. Sharma, and R. B. Gerber, Phys. Chem. Chem. Phys. **18**, 1607 (2016).
- ⁸³I. S. Ulusoy, Y. Scribano, D. M. Benoit, A. Tschetschetkin, N. Maurer, B. Koslowski, and P. Ziemann, Phys. Chem. Chem. Phys. **13**, 612 (2011).
- ⁸⁴S. K. Chulkov and D. M. Benoit, J. Chem. Phys. **139**, 214704 (2013).
- ⁸⁵D. M. Benoit, J. Phys. Chem. A **119**, 11583 (2015).

Appendix A: Derivation of Eq. (15)

For calculating the coefficients $p_{\mathbf{c}_l}^{\text{VCR}}$ of the cut function $F_{\mathbf{c}_l}$ in an incremental expansion for a variable combination range VCR according to Eq. (14) we need to consider all variable combinations \mathbf{c}_s in VCR that contain \mathbf{c}_l or are equal to \mathbf{c}_l , i.e. $\mathbf{c}_s \supseteq \mathbf{c}_l$. Since we require the VCR to be closed on forming subsets, the contribution of a superset \mathbf{c}_s to the coefficient of \mathbf{c}_l only depends on the order of the involved variable combinations, i.e., the values of l and s . They can be summarized as

$$\tilde{p}_l^s = \begin{cases} 0 & \forall s < l \\ 1 & \forall s = l \\ -\sum_{l'=l}^{s-1} \binom{s-l}{l'-l} \tilde{p}_l^{l'} & \forall s > l \end{cases}. \quad (\text{A1})$$

For variable combinations \mathbf{c}_s that are smaller than \mathbf{c}_l , \mathbf{c}_l cannot be a subset of or equal to \mathbf{c}_s and we obtain $\tilde{p}_l^s = 0 \quad \forall s < l$. For $s = l$, $F^{\mathbf{c}_l} = F^{\mathbf{c}_s}$ occurs only once in $\bar{F}^{\mathbf{c}_s}$ [see also Eq. (12)], so that $\tilde{p}_l^s = 1 \quad \forall s = l$. If $s > l$, $F^{\mathbf{c}_l}$ can contribute to several terms in $\bar{F}^{\mathbf{c}_s}$ [Eq. (12)]. It is contained in the bar potentials for all variable combinations that are

supersets of \mathbf{c}_l . Since these terms are correcting terms in the bar potential $\bar{F}^{\mathbf{c}_s}$, they have to be subtracted, which explains the minus sign of the contribution. Each term occurs $\binom{s-l}{l'-l}$ times, which is the number of subsets of \mathbf{c}_s that contain \mathbf{c}_l and have the order l' . Let us, for example look at the coefficient of the variable combination $\{A\}$. Its coefficient in $\bar{F}^{\{A\}}(\{x_A\})$ is trivially

$$\tilde{p}_1^1 = 1 \quad (\text{A2})$$

In

$$\bar{F}^{\{A,B\}}(\{x_A, x_B\}) = F^{\{A,B\}}(\{x_A, x_B\}) - \bar{F}^{\{A\}}(\{x_A\}) - \bar{F}^{\{B\}}(\{x_B\}) \quad (\text{A3})$$

$F^{\{A\}}(\{x_A\})$ has the coefficients

$$\tilde{p}_1^2 = -\tilde{p}_1^1 = -1 \quad (\text{A4})$$

and to

$$\begin{aligned} \bar{F}^{\{A,B,C\}}(\{x_A, x_B, x_C\}) &= F^{\{A,B,C\}}(\{x_A, x_B, x_C\}) \\ &\quad - \bar{F}^{\{A,B\}}(\{x_A, x_B\}) - \bar{F}^{\{A,C\}}(\{x_A, x_C\}) - \bar{F}^{\{B,C\}}(\{x_B, x_C\}) \\ &\quad - \bar{F}^{\{A\}}(\{x_A\}) - \bar{F}^{\{B\}}(\{x_B\}) - \bar{F}^{\{C\}}(\{x_C\}) \end{aligned} \quad (\text{A5})$$

it contributes with

$$\tilde{p}_1^3 = - \left(\binom{2}{0} \tilde{p}_1^1 + \binom{2}{1} \tilde{p}_1^2 \right) = -(1 \cdot 1 + 2 \cdot (-1)) = 1. \quad (\text{A6})$$

These observations can easily be generalized to

$$\tilde{p}_l^l = 1, \quad \tilde{p}_l^{l+1} = -1, \text{ and } \tilde{p}_l^{l+2} = 1. \quad (\text{A7})$$

This suggests that

$$\tilde{p}_l^s = (-1)^{s-l}. \quad (\text{A8})$$

We will prove in the following by induction that, if Eq. (A8) holds for $l < s < (n+1)$ (as shown above for $l < s < l+3$), it is also true for $s = n+1$ and thereby for any $s > l$,

$s, l \in \mathbb{N}$. Therefore we consider

$$\begin{aligned}
\tilde{p}_l^{n+1} &= - \sum_{l'=l}^n \binom{n-l+1}{l'-l} \tilde{p}_{l'}^{l'} = - \sum_{l'=l}^n \binom{n-l+1}{l'-l} (-1)^{l'-l} \\
&= - \sum_{l'=l}^{n+1} \binom{n-l+1}{l'-l} (-1)^{l'-l} + \binom{n-l+1}{n-l+1} (-1)^{n+1-l} \\
&= (-1)^{n+1-l},
\end{aligned} \tag{A9}$$

where we have used $\sum_{l''=0}^{n-l+1} \binom{n-l+1}{l''} (-1)^{l''} = 0$ with $l'' = l' - l$. Thereby we have proven that the coefficient of a variable combination \mathbf{c}_l to the bar function of a fragment combination $\mathbf{c}_s \supseteq \mathbf{c}_l$ is given by $\tilde{p}_l^s = (-1)^{s-l}$.

With this we can obtain the coefficients $p_{\mathbf{c}_{l'}}^{\text{VCR}}$ used in Eq. (14) as the sum of the coefficients $\tilde{p}_{l'}^l$ arising from all higher-order variable combinations \mathbf{c}_l in VCR that are supersets of $\mathbf{c}_{l'}$ as

$$p_{\mathbf{c}_{l'}}^{\text{VCR}} = \sum_{\mathbf{c}_l \in \text{VCR}; \mathbf{c}_l \supseteq \mathbf{c}_{l'}} \tilde{p}_{l'}^l = \sum_{\mathbf{c}_l \in \text{VCR}; \mathbf{c}_l \supseteq \mathbf{c}_{l'}} (-1)^{l-l'}. \tag{A10}$$

Thereby, we have derived Eq. (15).

Appendix B: Effective spatially restricted fragment combination range with only neighbor couplings for a chain-like system

Following Eq. (15), the coefficient of a particular fragment combination \mathbf{f}_k in the overall incremental expansion of the energy, $p_{\mathbf{f}_k}^{\text{FCR}}$, is dependent on the FCR. Assume, we have a chain-like system and a spatially restricted FCR with only neighbor couplings and a highest fragment-combination order of l . When adding a fragment combination of size $l+1$, we need additionally to add all subsets that are not yet included. For instance, the inclusion of a fragment combination $\{A, B, C, D\}$ to the $\text{FCR}^{\text{3F,NB}}$ in Eq. (25), has to be accompanied by the addition of the fragment combinations $\{A, D\}$, $\{A, B, D\}$, and $\{A, C, D\}$. This ensures that the resulting FCR is closed on forming subsets. All these additionally required fragment combinations contain the outer-most fragments of the new fragment combination, i.e., A and D in the present example. The number of required additional fragment combinations for a certain fragment combination level k is, hence, determined by the number of additional permutations possible. It can be obtained by $\binom{l'-2}{k-2}$, where $l' = l+1$ is the level of the newly added fragment combination $\mathbf{f}_{l'}$.

The extension of the FCR by $\mathbf{f}_{l'}$, ensuring that it is closed under forming subsets, may hence affect the coefficients in the expansion analogous to Eq. (14) for all fragment combinations that are subsets of or equal to $\mathbf{f}_{l'}$. We denote the change in coefficients for a fragment combination \mathbf{f}_k by $\Delta p_{\mathbf{f}_k}$. First we consider the fragment combinations \mathbf{f}_k that are subsets of or equal to the added fragment combination $\mathbf{f}_{l'}$ itself but not of any other *added* fragment combination. This is the only the case for $\mathbf{f}_{l'}$ itself and the connected mode combinations of size $k = l' - 1$ ($\{A, B, C\}$, and $\{B, C, D\}$ in the above example) as well as the connected fragment combination of size $k = l' - 2$ that does not contain the outer fragments ($\{B, C\}$ in the above example). In these cases, we obtain $\Delta p_{\mathbf{f}_k} = (-1)^{l'-k}$, which is the contribution due to the largest added fragment combination ($\mathbf{f}_{l'}$) alone. Smaller connected fragment combinations will be contained in the additionally added fragment combinations. In these cases and for the unconnected fragment combinations, we have to “count” how often the particular fragment combination is contained in the added fragment combinations. In these cases we obtain, when adding a fragment combination of level l' ($\mathbf{f}_{l'}$), a change of overall coefficient for a fragment combination \mathbf{f}_k , where $\mathbf{f}_k \subset \mathbf{f}_{l'}$, by applying Eq. (15),

$$\Delta p_{\mathbf{f}_k} = (-1)^{l'-k} + \sum_{i=k+2-o}^{l'-1} \binom{l'-k-2+o}{i-k-2+o} (-1)^{i-k}, \quad (\text{B1})$$

where o is the number of outer-most fragments of $\mathbf{f}_{l'}$ contained in \mathbf{f}_k , i.e., $o \in \{0, 1, 2\}$, so that $k \geq o$ and $k \leq l'$. The first term in Eq. (B1) arises from the added fragment combination $\mathbf{f}_{l'}$ and the sum collects all contributions due to the additional fragment combinations added to ensure the FCR to be closed on forming subsets. The binomial coefficient is obtained by consideration on how many permutations need to be considered when fixing the k fragments in \mathbf{f}_k and the $2 - o$ outer-most fragments that are not contained in \mathbf{f}_k . As an example, we consider $\Delta p_{\mathbf{f}_k}$ for $\mathbf{f}_k = \{A, B\}$, with $k = 2$, in the above example, where $\{A, B, C, D\}$ was added to the FCR^{3F,NB} in Eq. (25). The size of the largest added fragment combination is, hence, $l' = 4$. We see that $\{A, B\}$ contains one outer-most fragment, namely A, so that $o = 1$ and obtain

$$\Delta p_{\mathbf{f}_k} = (-1)^{4-2} + \sum_{i=3}^3 \binom{1}{0} (-1)^{3-2} = 1 - 1 = 0. \quad (\text{B2})$$

The first contribution, here is that of the added fragment combination $\{A, B, C, D\}$, and the second one that of $\{A, B, D\}$.

In the special case that $k = 0 \wedge o = 0 \wedge l' \leq 2$, the sum in Eq. (B1) does not contribute

TABLE VI. Coefficients p^{fF} and change of coefficient $\Delta p^{(f+1)F-fF}$ when adding the next $f+1$ -level of fragment combinations to the spatially restricted FCR with only neighbor couplings of a chain-like system with N fragments for different fragment combinations (FC) and fragment combination order f .

FC	p^{1F}	Δp^{2F-1F}	p^{2F}	Δp^{3F-2F}	p^{3F}	Δp^{4F-3F}	p^{4F}	...
$\{\}$	$(N-1)$	$(N-1) \cdot (-1)$	0	0	0	0	0	...
$\{F1\}, \{FN\}$	1	$1 \cdot (-1)$	0	0	0	0	0	...
$\{F2\}, \dots, \{F(N-1)\}$	1	$2 \cdot (-1)$	-1	$1 \cdot (1)$	0	0	0	...
$\{F1, F2\}, \{F(N-1), FN\}$	—	1	1	$1 \cdot (-1)$	0	0	0	...
$\{F2, F3\}, \dots, \{F(N-2), F(N-1)\}$	—	1	1	$2 \cdot (-1)$	-1	$1 \cdot (1)$	0	...
$\{F1, F2, F3\}, \{F(N-2), F(N-1), FN\}$	—	—	—	1	1	$1 \cdot (-1)$	0	...
$\{F2, F3, F4\}, \dots, \{F(N-3), F(N-2), F(N-1)\}$	—	—	—	1	1	$2 \cdot (-1)$	-1	...

and we obtain

$$\Delta p_{\mathbf{f}_k} = (-1)^{l'} \quad \forall (k = 0 \wedge o = 0 \wedge l' \leq 2). \quad (\text{B3})$$

In the other cases, we combine the sum with the contribution of the largest added fragment of size l' in Eq. (B1), so that

$$\Delta p_{\mathbf{f}_k} = \sum_{i=k+2-o}^{l'} \binom{l'-k-2+o}{i-k-2+o} (-1)^{i-k} = \sum_{i'=0}^{l''} \binom{l''}{i'} (-1)^{i'} = 0 \quad \forall \neg(k = 0 \wedge o = 0 \wedge l' \leq 2), \quad (\text{B4})$$

where $l'' = l' - k - 2 + o > 0$ and $i' = i - k - 2 + o$.

This means that the only fragment combinations that are subject to a non-zero change in the coefficients in this scheme are those of the added fragment combination of size l' itself as well as the connected fragment combination of one order lower than the added one ($k = l' - 1$), the connected fragment combination with $k = l' - 2$ that does not contain an outer-most fragment, and the zero set, if $l' \leq 2$. The contributions of all other cases cancel exactly. A consequence of this is that, for a chain-like molecule using a spatially restricted FCR with only neighbor couplings, there are no contributions from unconnected fragment combinations. This holds true for all levels of fragment combinations.

Building up the spatially restricted FCR with only neighbor couplings for a chain-like system is exercised in Table VI. We see that the first time a fragment combination of size l is introduced it has a coefficient of 1. The inclusion of the next-higher $(l+1)$ -level of fragment combinations then leads to two (or one in case of outer fragment combination) contributions

of -1 . This is due to the fact, that these fragment combinations are connected fragment combinations of level l that do not contain both outer fragments of a fragment combination of level $l + 1$. In case these l -level fragment combinations do not contain any of the outer-most fragments of the entire chain-like system, they are contained in two of the connected fragment combinations of size $l + 1$. This leads to a coefficient of $1 + 2 \cdot (-1) = -1$. If it contains one outer-most fragment of the chain, it is only contained in one of the added fragment combinations of size $l + 1$, leading to a coefficient of $1 + 1 \cdot (-1) = 0$. Adding the next level ($l+2$), can only change the coefficients by 1 for those l -level fragment combinations that are contained in an $(l + 2)$ -level fragment combination, but do not contain any outer fragment of the latter fragment combination. This case occurs exactly once for each l -level fragment combination that does not contain an outer-most fragment of the chain. The coefficients of these in the $(l + 2)F$ expansion are therefore zero. As shown above, the addition of higher-order fragment combinations in this scheme will not have any influence on the coefficients, so that they will be zero for all these higher FCRs. The results can be summarized for $f > 1$ as

$$p_{\mathbf{f}_k}^{fF,NB} = \begin{cases} 1 & \forall k = f \\ -1 & \forall k = (f - 1) \wedge \mathbf{f}_k \text{ connected} \wedge \mathbf{f}_k \cap \{F1, FN\} = \emptyset, \\ 0 & \text{else} \end{cases} \quad (\text{B5})$$

where $\{F1, FN\}$ is the set of the two outer-most fragments in the chain-like system. This means that the only fragment combinations that contribute to a spatially restricted f -level FCR in a chain like system are those of level f and $f - 1$, where the latter only have a non-zero contribution, in case they do not contain an outer-most fragment of the chain.

Appendix C: Coordinate transformation of polynomial PES representations

In the current work, we apply transformation between polynomial representations of PESs for certain fragment combinations (FCs). The starting auxiliary PES is represented in a polynomial sum-over-product form in auxiliary coordinates as,

$$V_{FC}^{\text{aux}}(\{q_{FC}^{\text{aux}}\}) = \sum_{\mathbf{t}} (a_{FC}^{\text{aux}})_{\mathbf{t}} \prod_{m \in \mathbf{t}} q_m^{\text{aux}}, \quad (\text{C1})$$

where \mathbf{t} denotes a particular term in the potential. For instance, $\mathbf{t} = \{q_1, q_1, q_1, q_3, q_3, q_5\}$ corresponds to the potential term $q_1^3 \cdot q_3^2 \cdot q_5$. Due to permutational symmetry, the same potential

term is also obtained for other \mathbf{t} s. This is the case, for instance, for $\mathbf{t}_2 = \{q_1, q_1, q_5, q_3, q_3, q_1\}$ and $\mathbf{t}_3 = \{q_1, q_1, q_5, q_3, q_1, q_3\}$ and we have $(a_{\text{FC}}^{\text{aux}})_{\mathbf{t}} = (a_{\text{FC}}^{\text{aux}})_{\mathbf{t}_2} = (a_{\text{FC}}^{\text{aux}})_{\mathbf{t}_3}$. We can, hence, collect these terms and express the PES as

$$V_{\text{FC}}^{\text{aux}}(\{q_{\text{FC}}^{\text{aux}}\}) = \sum_{\mathbf{s}} (c_{\text{FC}}^{\text{aux}})_{\mathbf{s}} \prod_{m \in \mathbf{s}} (q_m^{\text{aux}})^{e_m^{\mathbf{s}}}, \quad (\text{C2})$$

where \mathbf{s} contains the modes in this term as well as the power $e_m^{\mathbf{s}}$ with which it is represented in the potential term. For \mathbf{t} and \mathbf{s} corresponding to the same term, we have

$$(c_{\text{FC}}^{\text{aux}})_{\mathbf{s}} = u_{\mathbf{s}} (a_{\text{FC}}^{\text{aux}})_{\mathbf{t}}, \quad (\text{C3})$$

where $u_{\mathbf{s}} = \frac{(\sum_{m \in \mathbf{s}} e_m^{\mathbf{s}})!}{\prod e_m^{\mathbf{s}}!}$ is the number of possible permutations, i.e., the number of corresponding \mathbf{t} s.

Also the final PES in common FALCON coordinates can be written in both ways, i.e., as

$$V_{\text{FC}}^{\text{F}}(\{q_{\text{FC}}^{\text{F}}\}) = \sum_{\mathbf{t}} (a_{\text{FC}}^{\text{F}})_{\mathbf{t}} \prod_{m \in \mathbf{t}} q_m^{\text{F}}, \quad (\text{C4})$$

or

$$V_{\text{FC}}^{\text{F}}(\{q_{\text{FC}}^{\text{F}}\}) = \sum_{\mathbf{s}} (c_{\text{FC}}^{\text{F}})_{\mathbf{s}} \prod_{m \in \mathbf{s}} (q_m^{\text{F}})^{e_m^{\mathbf{s}}}, \quad (\text{C5})$$

where q_{FC}^{F} refers to that part of the common FALCON coordinates that displace the atoms in the respective fragment combination (FC).

The transformation from the PES in auxiliary coordinates $\{q\}_{\text{FC}}^{\text{aux}}$ to $\{q\}_{\text{FC}}^{\text{F}}$, can be performed via subsequent addition of terms for the $\{q\}_{\text{FC}}^{\text{F}}$ coordinates as linear combinations of PES terms containing auxiliary coordinates. The weights in this linear combination are given by the elements of the transformation matrix \mathbf{A}_{FC} . We can, for instance obtain,

$$(a_{\text{FC}}^{\text{F}})_{m_1, m_2, \dots, \tilde{m}_f, \dots, m_n} = \sum_{a \in \mathbf{m}^{\text{aux}}} (\mathbf{A}_{\text{FC}})_{fa} (a_{\text{FC}}^{\text{aux}})_{m_1, m_2, \dots, m_a, \dots, m_n}, \quad (\text{C6})$$

where \tilde{m}_f denotes the *new* common FALCON coordinate and a any auxiliary coordinate. When collecting the terms similar to above, we have to take the permutational symmetry into account,

$$(c_{\text{FC}}^{\text{F}})_{\tilde{\mathbf{s}}(f)} = u_{\tilde{\mathbf{s}}(f)} \sum_{a \in \mathbf{m}^{\text{aux}}} (\mathbf{A}_{\text{FC}})_{fa} u_{\tilde{\mathbf{s}}(a)}^{-1} (c_{\text{FC}}^{\text{aux}})_{\tilde{\mathbf{s}}(a)}, \quad (\text{C7})$$

where $\tilde{\mathbf{s}}(f)$ indicates, similar to above, that one coordinate of the original PES representation in auxiliary coordinates has been replaced by a new, common FALCON coordinate and in $\tilde{\mathbf{s}}(a)$ the same spot is occupied by the auxiliary coordinate a . \mathbf{m}^{aux} is the full set of all auxiliary coordinates. In this way we can add term after term for the new coordinates. The terms with higher order e_f in the new coordinates are subsequently obtained from the $e_f - 1$ terms that contain auxiliary coordinates.

In the transformation algorithm, the potential terms containing auxiliary coordinates are kept and re-used in subsequent addition of coordinates. Despite the fact that we transform an n -mode representation in auxiliary coordinates to an n -mode representation in common FALCON coordinates, coefficients for $(n + 1)$ -mode combinations with at least one auxiliary coordinate occur temporarily. We do not consider any terms of higher-order in mode combination, since they only contribute to $(n + 1)$ - or higher-order mode combinations in the final PES representation, which we neglect. The described subsequent addition of coordinates to a PES in two-mode representation in this way scales with M_{aux}^3 and M_{F}^2 , where M_{aux} is the number of auxiliary coordinates and M_{F} that of common FALCON coordinates to be considered in the respective fragment combination. At the end of the transformation, all contributions from auxiliary coordinates are removed from the representation of the PES.